Fall 2017

MICROBIAL SOURCE TRACKING IN COASTAL RECREATIONAL WATERS OF SOUTHERN MAINE: RELATIONSHIPS BETWEEN ENTEROCOCCI, ENVIRONMENTAL FACTORS, POTENTIAL PATHOGENS, AND FECAL SOURCES

Derek Rothenheber
University of New Hampshire, Durham

MICROBIAL SOURCE TRACKING IN COASTAL RECREATIONAL WATERS OF SOUTHERN MAINE: RELATIONSHIPS BETWEEN ENTEROCOCCI, ENVIRONMENTAL FACTORS, POTENTIAL PATHOGENS, AND FECAL SOURCES

BY

DEREK ROTHENHEBER

Microbiology BS, University of Maine, 2013

THESIS

Submitted to the University of New Hampshire

In Partial Fulfillment of

the Requirements for the Degree of

Master of Science

in

Microbiology

September, 2017

This thesis has been examined and approved in partial fulfillment of the requirements for the degree of Masters of Science in Microbiology by:

Thesis Director, Dr. Stephen H. Jones
Research Associate Professor
Department of Molecular, Cellular, and Biomedical Sciences

Dr. Jeffrey Foster
Assistant Professor
Department of Molecular, Cellular, and Biomedical Sciences

Dr. W. Kelley Thomas
Professor
Department of Molecular, Cellular, and Biomedical Sciences

Date

Original approval signatures are on file with the University of New Hampshire Graduate School

Table of Contents

ABSTRACT

MICROBIAL SOURCE TRACKING IN COASTAL RECREATIONAL WATERS OF SOUTHERN MAINE: RELATIONSHIPS BETWEEN ENTEROCOCCI, ENVIRONMENTAL FACTORS, POTENTIAL PATHOGENS, AND FECAL SOURCES

by

Derek Rothenheber

University of New Hampshire, September, 2017

Coastal water quality has been an issue for centuries. Fecal pollution of these waters represents a significant public health concern, as a variety of fecal sources can harbor human pathogens that range from bacteria and viruses to protozoa. Anthropogenic activities have exacerbated water quality through development such as housing, sewage infrastructure, and agriculture have increased fecal pollution sources and transport pathways, ultimately leading to increased pollution loading at the coast. For decades, federal, state and local municipalities have been using fecal indicator organisms to assess coastal water quality to assess public health risks. This thesis focuses on key factors that may influence enterococci concentrations, the fecal indicator for coastal recreational waters, and its relationship to fecal-borne bacterial pathogens to better understand its effectiveness as an indicator of public risk from fecal pollution. Results presented here come from three field sampling studies conducted during 2015–2016 where a combination of applied and basic research objectives were explored. Significant findings indicate that concentrations of enterococci are influenced predominantly by particle-associated enterococci and mammal fecal source concentrations across freshwater and estuary/marine environments. Other ecosystem-specific characteristics, such as sediment and freshwater transport, are also significant factors under some conditions. The relationship between fecal contamination and potential pathogens is water-type and location specific, with storm water having the highest

detection of fecal potential pathogens across a diverse data set. Moreover, bird feces represent a significant source of fecal-potential pathogens, as this fecal source significantly correlated to fecal potential pathogen abundance. Overall, the results highlight the dynamic nature of enterococci as a fecal indicator across different ecosystems. Ultimately, concentrations observed in the water are reflective of a combination of factors where potential health significance is location and water-type specific.

Chapter 1: Background and Literature Review

Introduction

Fecal pollution affects the quality and public health safety of surface waters across the United States. Fecal contamination can originate from human and non-human sources, including but not limited to inadequate wastewater treatment infrastructure, improper waste disposal, on-site septic system failure, wildlife, urban and agricultural runoff, pets, and livestock. The public health concern resides in the ability of animals, especially humans, to harbor a range of enteric pathogens that can make humans sick. These pathogens can include a wide array of bacterial, viral, or protozoan species, and the type or number of pathogens within surface waters at any time is difficult to measure. Thus, the direct measurement of specific pathogens to determine water quality is unfeasible as a basis for any regulatory guidelines so measurement alternatives have been developed.

Growing public concern and the need for a regulatory framework led to the passage of Clean Water Act (CWA) in 1972, which was an amendment to the Federal Water Pollution Control Act of 1948 (1). The CWA established a basic structure for regulating pollutant discharge and allowed for enforcement of water quality standards at the federal level, and it also detailed many other factors regarding water quality policy. For sewage pollution, water quality standards were developed around fecal indicator bacteria, and enterococci became established as the indicator used for coastal recreational waters. These indicators were established based on work published by Bonde (1966), who laid out criteria for an ideal indicator organism (2). Bonde stated that a fecal indicator needed to (a) be present whenever the pathogens of concern are, (b) occur in

much greater number than the pathogens, and (c) grow readily on relatively simple media (1). This has been the backbone of how water quality has been monitored and regulated for decades and is the foundation for the establishment of new indicator organisms. The implementation of enterococci as a fecal indicator for coastal recreational waters did not come into effect until 1986 after various epidemiological studies found a more rigorous relationship between enterococci and public health outcomes at marine beaches around the country than for other fecal indicators (1–4). With this information the EPA established a criterion of 104 CFUs/100mLs as the regulatory guideline for coastal recreational waters, which is based on a threshold of 30 predicted gastrointestinal (GI) illnesses for every 1000 swimmers (3), that is still used at present in New Hampshire, Maine and elsewhere around the US.

Of all the different sources of fecal contamination, human sources are considered the have greatest public health risk as the number of pathogens humans can carry is substantial and there is no "species barrier" as may occur with pathogens from other animals (8–10). Most coastal beaches, however, are affected by an array of largely unidentified sources considered to be non-point source pollution (7, 11). Non-point source contamination becomes elevated during heavy rainfall and run-off events and is often attributed to unidentified source types and locations until investigated to identify sources. Considering fecal indicator bacteria (FIB) are only indicators of general fecal contamination and are not source-specific, areas with elevated non-point source FIB levels are difficult to manage (12). If recreational waters do not meet the EPA and/or State criteria for water quality, as laid out in the Clean Water Act, the EPA or the State can enforce total maximum daily loads (TMDLs). This is a regulatory term that describes a plan for restoring impaired waters by identifying the maximum pollutant load a body of water can receive and still

meet water quality standards (13). This requires identification of pollutant sources, so the enforcement of such regulation helped develop the field of Microbial Source Tracking (MST) to enable accurate identification of non-point and specific fecal sources. MST has since been a key tool for supporting TMDL implementation (14).

Early pre-MST methods relied on the fecal coliform/fecal streptococcus ratio measured in water to suggest sources, where a ratio >4.0 represented human contamination and a ratio <0.7 indicated non-human pollution (15). The rationale for this was that human feces contains higher levels of fecal coliforms and non-human sources contain higher levels of fecal streptococci, however this method has since been abandoned as a viable approach due to highly variable and non-reproducible results (16). MST methods then shifted towards profiling fecal indicator bacteria via antibiotic resistance (17) and molecular characterization with pulse-field gel electrophoresis, repetitive element PCR, and ribotyping (18). These methods involved generating a library of cultured isolates from known fecal sources and then comparing fingerprints/profiles of isolates from an unknown water sample to the library. Both antibiotic resistance profiling and ribotyping were widely applied in the field from the late 1990s into the 2000s and were considered effective in tracking human and non-human fecal sources (19–21). Although antibiotic resistance profiling was rapid and could discriminate multiple different sources, there were significant limitations that ultimately led to its abandonment in mainstream MST applications. In particular, bacteria had to carry antibiotic resistance genes to be typed, a large geographically specific library/database was required, and in most cases antibiotic genes are carried on plasmids and often lost in environmental conditions (21). Out of all molecular characterization methods, ribotyping was considered the most effective (22). This method is,

however, extremely labor intensive, involving bacterial culture, DNA extraction, gel electrophoresis, restriction enzyme digest, Southern blotting, and analysis of DNA fingerprints (16). Not only is this method labor intensive, but the library of known sources needs to be region specific for this tool to be effective (20). Given all these limitations, ribotyping was never fully employed as a regulatory tool, however these limitations helped inform how the current MST methodologies were developed that are in use today.

1.1 Public Health Concern in Relation to Fecal Contamination

Fecal contamination represents a significant water quality and public health concern, as fecal material from various animals can harbor a wide variety of pathogens and significant levels of investment are needed to continuously address this issue (23). A study conducted in California estimated there are between 627,800 – 1,479,200 annual GI illnesses across 28 popular beaches in Southern CA, resulting in ~$21 million dollars annual economic loss (24). This study highlights not only the public health issue, but also the need for significant economic investment to manage water quality. Human pathogens vary widely in persistence, diversity and concentrations in contaminated waters and are often present at low concentrations, making detection of the full array very costly and logistically impossible (25). Fecal indicator bacteria are used instead of specific pathogen detection to provide an indication of the risk in a given waterbody relative to the levels detected. Human contamination is the greatest public health risk as humans represent the greatest potential to harbor pathogens that can make other humans sick. Surface waters are often contaminated by other sources of fecal contamination, however, which also pose human health risks. Below is brief overview of the relative risk that major types of fecal sources represent to human health.

1.1.1 Human/Sewage Fecal Source

The World Health Organization (WHO) reports that human sewage can contain pathogenic bacteria, viruses, parasitic protozoa, and helminths (26). The most common illness associated with contaminated waters is GI illness, as that is the main mode of action for these pathogens transmitted via the fecal-oral route. From 1985–1998 the WHO reported the main etiological agents associated with GI illness outbreaks in recreational waters were *Shigella* spp., *E. coli* 0157:H7, and *Giardia lamblia* (26). Since then, other studies have shown that human viruses (i.e. Norovirus) may be the main etiological agent in GI illness associated with sewage contamination of recreational waters (8). Water bodies contaminated by human fecal/sewage sources are also likely to be chronic and non-intermittent, because most common sources of human contamination are from faulty infrastructure, thus maintaining potential exposure to pathogens (9). These sources include but are not limited to, storm water (27), wastewater effluent (28, 29), malfunctioning septic systems (30), and shedding from humans within the water (31). These functional sources of human fecal pollution can, in theory, be eliminated with adequate resource investment, in contrast to some of the non-human sources. Finally, although human fecal contamination poses the greatest health risk, pollution from other animal source can be significant, and non-human fecal pollution can have the potential to carry a variety of zoonotic pathogens.

1.1.2 Bird Fecal Source

Coastal recreational waters harbor large populations of gulls species, and many coastal beach managers associate poor water quality with gull populations (32). Gulls can harbor *Salmonella* and *Camplyobacter* spp. and their presence in coastal recreation waters has been specifically

linked to gull fecal contamination (33, 34). Beaches that harbor large populations of gulls represent a potential health risk, as these birds can be a chronic fecal input source. Similar to human sources, this type of fecal pollution source can be successfully managed. A study in Wisconsin showed removal of gulls via trained border collies significantly reduced the detection of *Salmonella* and *Campylobacter* spp. and significantly reduced the number of water quality advisories posted (34). Lethal (shooting) and non-lethal management approaches were used in Lewiston, Maine to eliminate gulls as sources of fecal pollution to a water supply and successfully reduce fecal coliform levels (35). There are, however, other issues with bird-borne fecal pollution. A study conducted by the Jones lab (2008) in New Hampshire showed that gulls could be significant transporters of fecal material from wastewater treatment plants to marine ecosystems (36). This study showed significant similarity between ribotypes of *E. coli* isolates from wastewater treatment facilities and gull guano recovered from an offshore island, suggesting herring gulls were transporting and depositing land-based human fecal material to the offshore islands. Some of the banded gulls were even observed as far south as Florida, suggesting gull fecal impact can be widespread (36).

1.1.3 Dog Fecal Source

Dogs are common household pets in the United States, and are also frequent companions to humans when they visit recreational waters. Thus, dogs can represent a significant source of fecal contamination at beaches, which is mostly the result of improper disposal of their waste in public areas. Recent studies have shown that dogs may act as a reservoir for antibiotic resistant bacteria, and can harbor a significant amount of *Giardia* cysts (37, 38). Dogs have also been shown to carry vancomycin resistant (VRE) enterococci, which is an emergent public health risk

(39). For example, a study showed 8% of dog fecal samples in a farm-setting contained VRE enterococci, which is likely a reflection of VRE use in livestock (40). Other studies have shown that 39% of human pathogens can also infect domestic dogs, an indication that dogs have the potential to be significant reservoirs for human pathogens (41). Theoretically, fecal-borne pollution from dogs can be easily managed via public education efforts and other related strategies (42).

1.1.4 Other Wildlife

Fecal contamination from wildlife also represents a public health risk to recreational water because of the wide range of zoonotic diseases that could be carried by diverse species. Although the greatest wildlife source of fecal contamination at a coastal recreational beach is likely to be birds/gulls (43), other wildlife could also be present. The likely distance of wildlife from the water's edge due to the high human presence suggests the probability of direct fecal deposition by wild animals and their public health significance is probably relatively low for coastal recreational waters. However, during heavy rain events (>1 in.) their fecal contamination can be transported to the coast via runoff, thus elevating their pollution and potential public health significance. Maine has the largest moose and black bear population in the "lower 48", however in Southern Maine, the presence of white-tailed deer, raccoons, muskrats, squirrels, foxes, and other small mammals are the likely wildlife fecal sources (44). Species of *Leptospira* can infect a wide range of wildlife hosts, and can be transmitted to humans through infective urine (45). Outbreaks are more common in recreational lakes and ponds, as transport of the pathogen to the coast at an infective dose is less likely (45). Rodents can carry Hantavirus that can be transmitted through fecal droppings, but this disease is far more prevalent in the Western United States and

only one case has been reported in Maine (Maine CDC, 2011). The most likely etiological agent for human infection from wildlife fecal source in this area is more likely to be a pathogenic bacterium (*Clostridium perfringens* or *E .coli*) (46). Non-bird wildlife sources are also the most difficult type of pollution to manage compared to other pollution source types, making remediation efforts problematic.

1.2 Enterococci in the Environment

Contamination of recreational waters based on fecal indicator organism concentrations represents potential fecal-borne public health risks to humans who swim and engage in other types of recreation where they can be exposed to contaminated water (13). For marine beach waters, there is a significant relationship between enterococci and negative health outcomes based on several epidemiological studies (1–4). It is important to note in relation to this study that these studies were conducted on beaches that had a significant point source of human/sewage contamination affecting the water quality. Although enterococci are used as an indicator of fecal contamination, they can also thrive in a wide range of habitats and are thus not exclusive to the fecal tract (47). Enterococci can persist in a variety of environments including fresh water sediments (48–50), marine sediments (48, 51), and soils (56–59). These non-fecal sources represent potentially significant non-point sources that can elevate enterococci levels measured in coastal recreational waters and skew the implied public health significance (47).

1.2.1 Enterococci in Soils

Early research into enterococci persistence in soils focused on anthropogenic activities, in particular cattle grazing locations (56). Results showed that agricultural activities significantly influenced runoff water quality and were the main pollution source of waters in the area (56, 57).

These soils contained elevated levels of fecal indicators as a result of manure application, however more recent research has shown that enterococci can also exist in similarly elevated concentrations in more naturalized soils (53, 55). A recent study conducted in Hawaii showed that enterococci presence was ubiquitous in surveyed soils, and concentrations were often over 1,000 MPN/g soil. Other studies have shown that enterococci can persist longer than *E. coli* in soils when exposed to environmental stresses, such as desiccation. One such study used mesocosms where soil (35% moisture) was seeded with *Enterococcus faecalis* and *E. coli* and allowed to desiccate (12% moisture) for 8 days. Results showed *E. faecalis* levels remained constant (10^6 CFU/g) whereas *E. coli* densities dropped to <1 CFU/g of soil in 4 days (58). The original source of enterococci in soils can differ from site to site, however it is clear that soils likely provide the necessary nutrients for subpopulations of enterococci to become naturalized (47, 55).

1.2.2 Enterococci in Sediments

Freshwater and tidal streams are major conduits for bacterial transport from watersheds to the coast. Fecal contamination of these streams results in increased fecal indicator bacteria loads. For these bacteria to reach coastal recreational waters, however, stream flow needs to be fast enough to support transport versus re-deposition. Studies that investigated underlying freshwater and marine sediments have found that enterococci concentrations can be orders of magnitude higher than the overlaying freshwater (59, 60) and marine water (59, 61). These results suggest that sediments can act as a major reservoir for enterococci that is the result of sedimentation of bacteria from the overlaying water (47) and ensuing persistence in this environment. Other studies have also shown that enterococci are able to persist longer in both freshwater sediments

(50) and marine sediments (51) when compared to their overlaying waters. As an environment, sediments are less likely to be influenced by environmental stressors (UV damage and starvation) and predation from bacteriovores, and thus might be more favorable for enterococci to persist and regrow (47). One study tracked *E. coli* populations through streambeds and focused on how sediment resuspension can influence concentrations in the water (62). They seeded a streambed with nalidixic acid resistant *E. coli* and tracked its persistence and resuspension during storm events. Results showed that during the rising limb of the hydrograph the nalidixic acid resistant *E. coli* significantly increased in the overlaying water (62). No such study has been conducted for enterococci. A more recent study that analyzed *E. coli* and enterococci resuspension in stream beds using Rep-PCR fingerprinting showed that resuspension of sediment-associated bacteria is not exclusive to high flow events and that sediments act as a virtually endless supply of fecal indicator bacteria (63).

Collectively, soil and sediments act as significant non-point sources of enterococci that can affect downstream recreational water quality measurements. The public health significance is likely lower than recreational waters contaminated by a point source of fecal pollution, but there are limited studies on this subject. One study showed that environmental sources of enterococci (eel grass beds) represent a significantly lower predicted risk to humans when compared to human and animal contamination (9).

1.3 Current Microbial Source Tracking Methods.

Many early MST methods, based on both phenotypic and molecular characterization, relied on extensive libraries of know sources to compare against unknown samples. Depending on the size of the database, an extensive number of known sources could be determined. The Jones lab at UNH in particular had the ability to determine 37 different fecal sources via ribotyping based on thousands of *E. coli* isolate patterns (64). Although these methods are robust, the requirement of an extensive regional database is a challenge for large-scale applications (65). Given this and other limitations, the MST field migrated towards library-independent methods. These methods target alternative culture-independent detection of organisms that are considered more exclusive to the fecal tract than *E. coli* or enterococci, and utilize rapid detection methods. Presented below are the major current library-independent MST assays that are the centerpiece for this thesis work.

1.3.1 Mammal and Human Fecal Assays

In the 2000's, work conducted by Bernhard and Field (66, 67) sought to design PCR assays that could target specific fecal sources. These researchers targeted the *Bacteroides* group of bacteria as an earlier study published by Kreader (1995), showed *B. distasonis, B. thetaiotaomicron, and B. vulgatus* were relatively specific to human fecal material when detected with PCR assays (68). *Bacteroides* spp. are a group of Gram-negative, non-spore forming, obligate anaerobes that are highly abundant in the intestine tracts of mammals. These organisms are also strict inhabitants of fecal tracts, so their presence in the environment is thought to be directly related to recent fecal contamination (68). This rationale led to the design of primers that target the 16S rRNA gene of all *Bacteroides* using a marker: Bac32, which is now commonly used as a marker for mammalian

fecal contamination (67). At the time, *Bacteroides* spp. composition in different mammals was thought to be unique, mostly because diets can vary drastically between species (69). The Bernhard and Field research group used length-heterogeneity PCR (LH-PCR) and terminal restriction fragment length polymorphism (T-RFLP) to characterize 13 healthy human fecal, 3 sewage, and 19 cow fecal samples using primers specific to the *Bacteroides* spp. 16S rRNA gene sequence (67). Clones that showed host specificity were cloned into a plasmid and Sanger sequenced. A Basic Local Alignment Search Tool (BLAST)

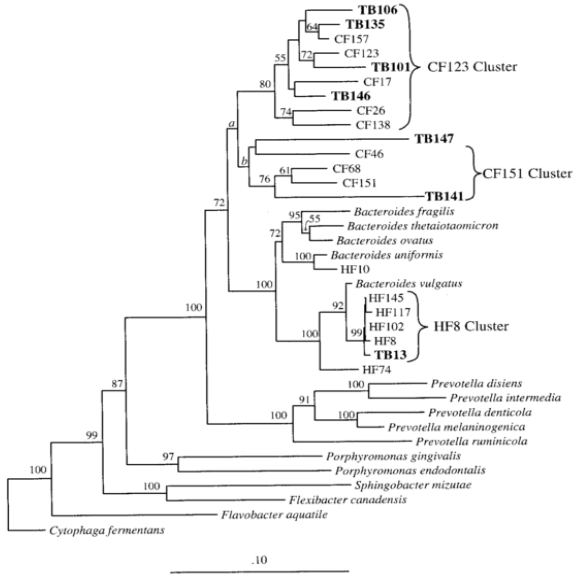


FIG. 1. Phylogenetic relationships among partial 16S rDNA sequences (558 positions) of clones recovered from Tillamook Bay water samples (TB). HF and CF are host-specific genetic markers identified from human and cow fecal clone libraries, respectively. The tree was inferred by neighbor joining. Numbers above the internal branches are percentages of bootstrap replicates that support the branching order. Bootstrap values below 50% are not shown. Bootstrap values for branches *a* and *b* dropped from 68 to 47 and 76 to 40, respectively, when TB147 was added to the analysis. The sequence from *Cytophaga fermentans* was used to root the tree.

Figure 1. Phylogenic relationships of *Bacteroides* spp. 16S rRNA. Results show human specific (HF8) and two ruminant specific (CF123 and CF151) clusters (67).

was then used to determine the relationship of isolates to known *Bacteroides* spp. that were in the GenBank database at the time. Results showed human (HF8) and ruminant specific (CF123 & CF151) clusters were distinct, with isolates being >98.9% similar in the HF8 cluster (Figure 1). Bernhard and Field later published primer sets that targeted the HF8 cluster (HF183 assay)

for humans and both CF123 and CF151 (CF128 and CF193 assays) clusters that targeted ruminants (66).

Human Assay (HF183) Evaluation.

Since its inception in 2000, the HF183 assay targeting human fecal source has been the first and longest-lived MST PCR assay. The main reason for its long-term use and rapid application in the field is because of its performance in various benchmarking studies. In 2009, Ahmed et al. (70, 71) published two evaluation studies of the HF183 assay. In the first paper (70), the HF183 assay was compared to three other human specific assays that were tested for specificity and sensitivity against five animal hosts (dogs, ducks, cattle, pigs, and sheep). The other three assays showed 100% specificity to human samples, while HF183 was the only assay to show cross reactivity because one dog sample screened positive, so HF183 was 98% specific to human feces. Serial dilutions of sewage-spiked samples with freshwater, seawater, and distilled water, however, showed HF183 to be highly specific to the sewage compared to the other assays. Ahmed's other paper (71) assessed four different assays specific for human fecal sources against the HF183 assay. The PCR assays were tested for specificity against 186 different fecal samples from a range of sources (wastewater influent, cattle, pigs, sheep, goat, horses, chickens, dogs, ducks, pelicans, and kangaroos). HF183 outperformed other assays, and was shown to be 99% specific to human sources. There have been other evaluation studies that have come to the same consensus. Harwood et al (2014) indicated that the HF183 assay was the best human-specific MST PCR currently available (25).

qPCR HF183 Assay

After the development of the HF183 endpoint assay, research groups were quick to develop qPCR assays. The first assay used SYBR Green chemistry and a new reverse primer design to shorten the product size from 551 bp to 86 bp (72). This assay has since been deemed non-optimal as it is not source specific (25). Later TaqMan assays were developed, and of importance here is the HF183/BFDRev assay (73). This assay has a product size of 167 bp, uses the same HF183 forward primer as the endpoint assay and contains a FAM labeled probe that sits 6 bps upstream of the reverse primer BFDRev. At the time this assay was considered the best, as it had been well evaluated and used in the field (25). However, Green et al. (2014), redesigned the reverse primer to reduce potential primer-dimer formations, and since then the redesigned assay has continued to perform well (74). For all work conducted for this thesis, the HF183/BFDRev qPCR assay was used to detect human fecal contamination.

1.3.2 Dog Fecal Assay

After the development of the human fecal source assay (HF183), many other research groups started to develop other fecal-source specific assays. Development of a dog specific fecal source assay proved to be difficult, as the population structures of *Bacteroides* spp. are very similar between dogs and humans. The development of the dog specific assay (DF475) used subtractive hybridization to enrich for target host-specific fecal *Bacteroides,* as earlier work conducted by Dick et al. (2005) showed *Bacteroides* clone libraries were very similar between humans, cats, and dogs (75). The authors used a pooled human fecal clone library as the source of subtracted sequences, which were then presented to a pooled dog clone library. Similar sequences hybridized to one another and were immobilized, then unique (non-hybridized) sequences were amplified to increase the abundance of unique sequences. The pool of unique sequences

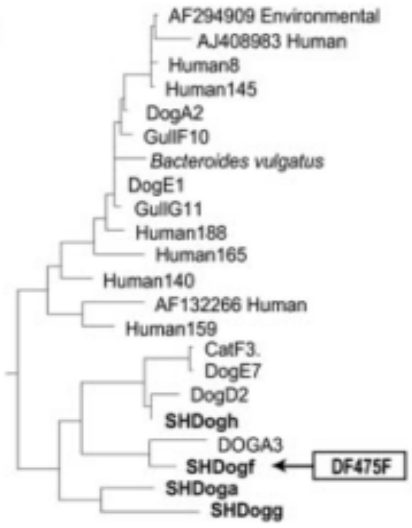


Figure 2. Dog Target Fecal Sequences Derived from Subtractive Hybridization. Results show derived dog sequences in relationship to non-dog sequences. The SHDogf was the fragment of choice for design of the DF475 assay (75).

revealed two dog-specific clades, and an assay was then developed based on the SHDogf sequence (Figure 2). The assay did not amplify fecal DNA from human, cat, cow, pig, chicken, or gull source, and has a sensitivity of 100 copies.

Dog Assay (DF475) Evaluation

To date there have been two major laboratory evaluations on dog feces source assays, however none have been conducted on the endpoint DF475 assay. Only BacCan-UCD and DogBact have been evaluated, both of which are qPCR assays. The DogBact assay uses the same primer set as the DF475 endpoint assay, but also contains a TaqMan probe that sits upstream of the reverse primer (76). A study conducted by Schriewer et al. (2013) evaluated both BacCan-UCD and DogBact and showed both assays exhibited 100% sensitivity, but both assays also had random cross-reactivity (77). The authors mentioned that a normalization process helped improve specificity of both assays to 83% for DogBact and 68% for BacCan-UCD. In the end the authors noted both assays performed equally well, even though there was a significant difference in

specificity between the assays (77). A different evaluation reported both assays showed cross-reactivity with non-target fecal source, but both were >80% sensitive and specific (78).

1.3.3 Gull/Bird Fecal Assay

Coastal recreational waters in the US are well populated by several species of gulls. They have even been observed to travel hundreds of kilometers inland for roosting grounds in Massachusetts and flock to inland wastewater treatment facilities (36, 43), making their presence and their capacity for transporting human pathogens to marine waters widespread (79). Lu et al. (2008) analyzed the bacterial community composition of gull feces obtained from West Virginia. The authors constructed a 16S rRNA gene clone library containing 282 sequences, and found the composition of *Bacilli* (37%), *Clostridia* (17%), *Gammaproteobacteria* (11%), and *Bacteroidetes* (1%). Furthermore, of the *Bacilli,* 26% were related to *Catellicoccus marimammalium* (80).

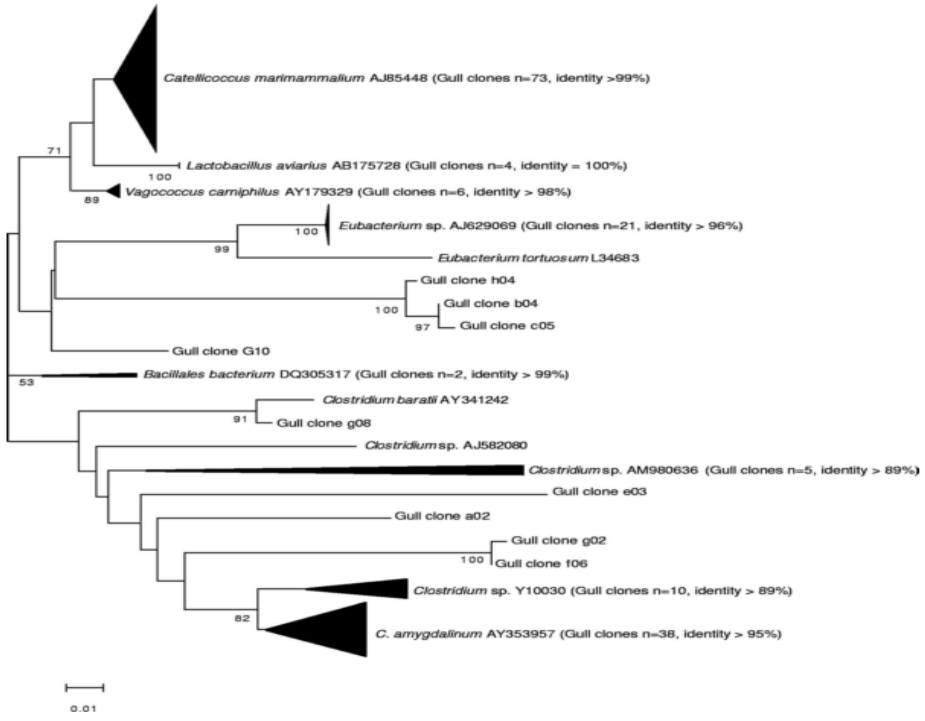


Figure 3. Unrooted neighbor-joining tree of 16S rRNA gene sequences from gull fecal clone libraries (80).

These results indicated that the gull gut is not dominated by *Bacteroides* spp., but rather low G+C genome content Gram-positive bacteria that are closely related to *C. marimammalium* (Figure 3). This organism made an easy target for source specific PCR assay development as it was highly abundant and unique when compared to mammals. The authors showed their PCR assay (Gull2) could detect fecal signatures from multiple gull species (*Larus domesticus* and *L. atricilla*) across multiple regions in the US and Canada, and that it was specific to gulls (i.e. no cross reactivity) (80).

Gull Assay (Gull2) Evaluation

Evaluation studies of the Gull2 marker have generally reported reliable performance in assay sensitivity and specificity. Boehm et al. (2013) showed that Gull2 was >80% sensitive and specific, with cross reactively coming from goose and pigeon fecal samples (78). A different study that also evaluated the Gull2 assay reported similar characteristics in assay sensitivity and specificity (81). The significance of this cross reactivity to goose and pigeons might not be of great significance depending on where the assay is applied, but also remediation strategies for avian fecal sources are likely similar (43, 81).

Bird qPCR Assay

The need to detect a range of avian sources is important for some field studies, as coastal watersheds in New England can be impacted by gulls, cormorants, turkeys and Canada geese, among others (82, 83). Previous studies have also shown that bird microbiomes are similar

across some avian species (gulls, chickens, and ducks) (80), and that other avian species cross react with assays that target gull fecal source (30, 33). Green et al. (2012) used a subtractive hybridization method to enrich for avian specific signatures to design an assay. This was a similar method used by Dick et al. (2005) in the development of the DF475 dog fecal source assay, but they used subtracter sequences from human, dog, cat, cow, and pig feces to enrich. Derived sequences mostly came from *Enterobacteriaceae* (47%) and *Helicobacter* (26%), which lead to the developed of the GFD assay that targeted an unclassified *Helicobacter* species (84). This assay was 100% specific to avian fecal source. Ahmed et al. (2016) also showed this assay to be robust in the detection of avian fecal sources (85).

1.3.4 Microbial Source Tracking with Next Generation Sequencing.

The advent of next generation sequencing (NGS) has helped almost every field in the biological sciences improve traditional analyses and develop new more innovative approaches. The MST field has certainly been no exception, as NGS has enabled development of new approaches to source tracking as well as survey approaches to tracking potential pathogen abundance (86). The first use of NGS technologies for MST were employed to evaluate the differences in microbiomes of various organisms for new marker development (87). A recent paper by Newton et al. (2013) used NGS approaches to identify signatures associated with sewer and fecal pollution (88). Their findings identified *Acinetobacter, Arcobacter, and Trichoccus* genera to be associated with sewer infrastructure, and the *Bacteroidaceae, Porphyomonadaceae, Clostridiaceae, Lachnospiraceae, and Ruminococcaceae* families to be associated with fecal pollution. This was one the first studies that used a "fingerprinting" strategy to identify unique taxa associated with different fecal contamination environments.

One of the more revolutionary developments has come from software published by Knights et al. (2012), called SourceTracker (89). This program uses 16S rRNA sequence data from Illumina platforms and can "fingerprint" bacterial communities from a known source and determine the proportion of each source within an unknown sample (Figure 4; 38).

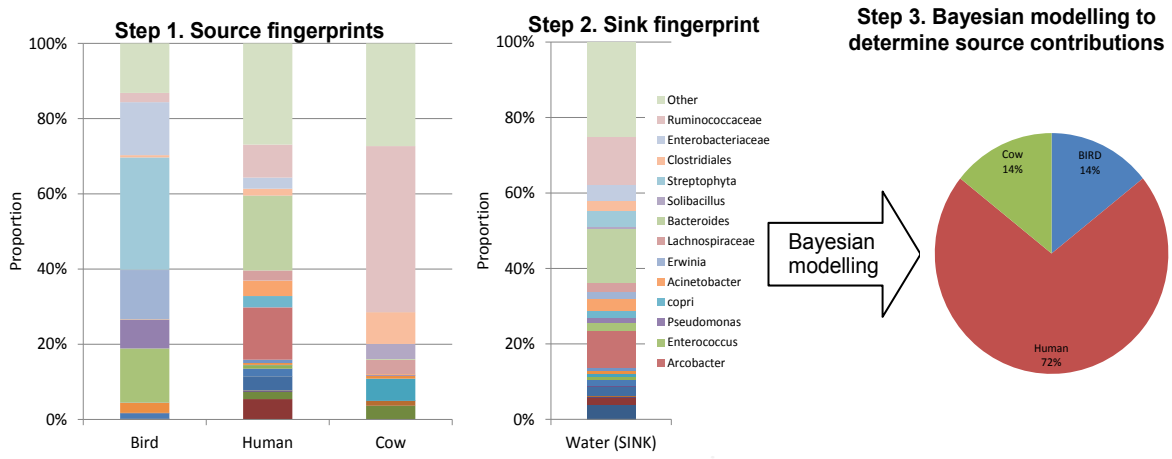


Figure 4. Visual Representation of SourceTracker Work Flow. SourceTracker comprises of three fundamental steps. Source samples and sink samples are fingerprinted (steps 1 & 2), then a Bayesian mixing model is used to determine sources within a sink (step 3). Figure from Henry et al. (90).

The program uses a Bayesian mixing model that employs a Gibbs sampling procedure for fingerprinting and assignment. In a general sense, SourceTracker is a robust version of ribotyping, in that a known library of sources needs to be generated for comparison with unknown samples. The original paper validated the method against a Random Forest and Naïve Bayes approaches by tracking gut, oral, skin, and soil sources in a laboratory and office environment (89). The authors showed that the method could discern different sources that were highly identical (as measured by the Jensen-Shannon divergence), and that SourceTracker was far more robust in assignments compared to other Bayesian approaches (89). To date, this tool

has been used for a wide variety of applications from contaminant tracking in neonatal units (91), to bacterial community mixing dynamics (92), and even more traditional MST approaches (88). The method was also evaluated in 2016 by a group in Australia, who reported that SourceTracker assigned sources that were >1% of an unknown sample and that default parameters were optimal in source assignment (90).

New research investigations with NGS technologies have employed the power of environmental DNA (eDNA), which has mostly been used for population surveys of various animals in the past (93, 94). The premise of eDNA relies on the assumption that an organism sheds its DNA as it interacts with the environment, studies have also shown that an organism's eDNA predominately exists within mitochondria (95). Barcoded primers can be used to amplify the DNA within a sample, which then can be analyzed in a similar fashion to bacterial 16S data. Specifically for microbial source tracking, a handful of assays that target humans, pigs, and cows have been designed to amplify mitochondrial DNA (96), these assays have however always underperformed their *Bacteroides* counterparts. To date, eDNA analysis has not been fully used in the microbial source tracking field, but it's incorporation into studies could provide valuable ecological context to a specific field site, refined geographic location of contamination sources, and the development of novel MST assays.

1.3.5 Persistence of MST DNA Markers in the Environment.

Current microbial source tracking methods rely on a DNA-based detection method. This detection strategy could either be detecting living or dead bacteria, and thus its persistence in the environment is imperative for interpretation of results. Mesocosms studies have shown that

under normal conditions detectable *Bacteroides* fecal markers persist between 7-10 days (97). Unlike enterococci, *Bacteroides* spp. cannot grow in aquatic ecosystems so concentrations observed within water are reflective of recent (within one week) fecal contamination. However, there are other factors that can influence this decay rate, especially UV exposure and temperature (97–99). Increased UV exposure and water temperatures can both significantly increase the DNA decay rate, resulting in the DNA marker persisting for much less than a week during summer months. This leads to concentrations of MST markers observed during the summer being more reflective of "real-time" fecal contamination and actually aids the overall interpretation of specific copy number concentrations. Overall, concentrations of MST markers observed within the environment do reflect recent contamination events, which can help interpret FIB concentrations as well.

1.4 How Microbial Source Tracking Can Inform Management Decisions.

Since the passage of the Clean Water Act in 1972, municipalities have been tasked with maintaining public recreational water quality across the country. For areas where the water quality does not meet standards that support designated uses, development of total maximum daily load model (TMDL) is required to direct required remediation efforts. This process involves calculating the maximum pollution load a water body can receive and still meet water quality standards. In most cases, determining point sources of fecal indicator bacteria results in significant reduction of pollution loads and improved water quality. In the past, much of this work, at the municipality level has involved targeted sampling strategies to hone in on "hot-spots" of contamination. However, using only detection of enterococci concentrations, these strategies have often been unsuccessful in part because enterococci are not fecal or

environmental source specific, so direct identification of a source remains difficult. This also made remediation efforts and TMDL required pollution reductions challenging and almost impossible for local towns to implement.

1.4.1 Simple Models to Inform Decision Making.

Not only are enterococci a non-source specific indicator for fecal contamination, it also takes at least 24 hours for a MPN or membrane filtration assay to provide results. This delay results in water sample concentrations being determined one day after collection, and thus not reflective of the actual current water quality. In the State of Maine, when coastal recreational beaches test positive for elevated levels of enterococci (>104 MPN/100 ml) an advisory is posted and held in place until follow-up sampling results in lower concentrations. Thus, not only does it take 24 hours to determine enterococci concentrations in water samples, but current regulations and follow-up advisories are always posted at least a day late.

In 2014, the Town of York, Maine wanted to create a proactive management tool that would allow them to predict when concentrations of enterococci are most likely to be elevated. This resulted in the Jones lab conducting an intensive field sampling campaign during the summer of 2014 to determine under what environmental conditions are enterococci detected at elevated concentrations in the beach water (100). Over the course of the season, samples were collected at least 5 days per week from July 3 to September 17, 2014, along with a variety of environmental variables and climatic conditions. Results identified two relatively significant types of point sources for enterococci, storm drains, especially during and following rainfall/runoff events, and

mounds of decaying seaweed (100). Enterococci concentrations in storm drains were shown to be above regulatory limits >69% of the time, and decaying seaweed mounds could contain levels >24,196 MPN/100 ml, the detection limit for the MPN assay used. Enterococci concentrations were also significantly related to rainfall amount, salinity, and tidal height (100). From all the 2014 data and previous years' data from the Maine Healthy Beaches Program, a rainfall advisory was proposed where beach advisories specific to different beach management areas should be posted if there is >1.5 inches of rain within 48 hours. The Town of York now uses this tool to post real-time beach advisories that are in effect during rainfall and a few other conditions where elevated concentrations of enterococci are most likely to occur. A similar approach was taken by the Maine Healthy Beaches program where rainfall advisories were developed for 20 different beaches affected by non-point source pollution (101). Other research groups have developed predictive models for regions in the USA, and their findings are similar in that runoff is identified as the most significant source of enterococci, leading to similar rainfall advisories constructed for use by local towns (102).

1.4.2 Identification of Fecal Hotspots

Given that enterococci are not a source-specific fecal indicator, remediation efforts based on enterococci alone is very difficult. The first step almost every town takes in in remediation efforts is to conduct a sanitary survey. Essentially, this is when field investigations are conducted to identify potential fecal sources within a watershed or coastal area. These surveys can be comprehensive, and can provide highly useful information and an understanding of animal species diversity, wastewater infrastructure, and potential pollutant pathways to recreational waters. Ultimately, these surveys lead to targeted-sampling approaches that seek to identify point

sources of pollution. Although these approaches are good first steps, they often result in expensive and long-term sampling efforts that can't directly identify fecal sources.

In the early 2000s, municipalities and water quality management agencies recruited the Jones lab to conduct MST studies using ribotyping. This was the first opportunity for the New England region to identify different types of fecal sources that were present in water samples. The application of ribotyping identified a wide range of predominant fecal sources across a wide range of studies, from feral cat contamination in two Southern Maine watersheds (44), human fecal contamination in Hampton Harbor, NH (103), to birds and wild animal fecal sources at New Castle Beach, NH (104). The results from these reports led to numerous follow-up investigations as well as infrastructure improvements that ultimately resulted in better management of the local water quality issues.

Although ribotyping can provide valuable information on fecal sources that are present, it is relatively expensive and time consuming. Hence, starting with the development of PCR based MST assays in the late 2000s, research laboratories and regulatory agencies have methodically focused efforts to evaluate these less expensive and time consuming methods for application and use in the field for identifying fecal contamination sources. These assays can be implemented at a fraction of the cost and time needed for ribotyping, and with the development of quantitative PCR assays, relative source strength can also be determined. Since 2015, the Jones lab has been employing the use of these MST PCR-based assays in conjunction with towns throughout the New England area. We have used the methods to identify spatial and temporal patterns in

targeted fecal sources, as well as to better locate fecal contamination hotspots. Specifically, in the Town of York, ME we identified storm drains that contained elevated levels of mammal, human, and bird fecal sources and we determined that the presence of these types of fecal sources is seasonally dependent (105). In the Town of Waldoboro, ME we showed that sites containing human fecal marker concentrations of $>10^5$ copies/100 ml are close to a source of human fecal contamination, like a sewer line break or illicit discharge (106). We also partnered with the Maine Healthy Beaches program in 2016 to identify fecal sources present in a historically polluted tidal river and marsh area located in the watershed of Old Orchard Beach. Our findings identified a hotspot of human fecal contamination that contained elevated levels of human fecal pollution during July and August ($>10^5$ copies/100 ml), and ruled out human fecal contamination as a significant source in marsh sites. This information has since been presented to the town, where follow-up infrastructure investigations and a new MST study are being conducted in 2017.

1.4.3 Using Regional Microbial Source Tracking qPCR Data to Help Convey the Significance of Fecal Pollution to the Public.

The use of fecal indicator bacteria has long been established and used to evaluate recreational water quality. Numerous studies have been conducted to evaluate the public health significance of fecal indicator bacteria, and have been the basis for establishing regulatory limits (13). For MST markers, there are no established limits for acceptable levels of contamination from various sources. Recently, a study designed to address this (Boehm et al. 2015), used a risk-based approach to show that <4,200 copies of HF183/100 ml is an acceptable threshold level of human fecal contamination (107). This limit is based on the EPA benchmark of 30 predicted GI illnesses/1000 swimmers, and is the first time a limit was determined for a MST marker. However, this limit is not part of any regulatory framework, and other MST markers don't have a

risk-based assessment. This makes conveying MST results especially challenging, as there is generally no context associated with measured concentrations and marker copy numbers. To address this issue, we used all the source-specific marker data collected in the region to compare study-to-study results, as well as to compare similar aquatic habitats (e.g. marshes, freshwater sites.)

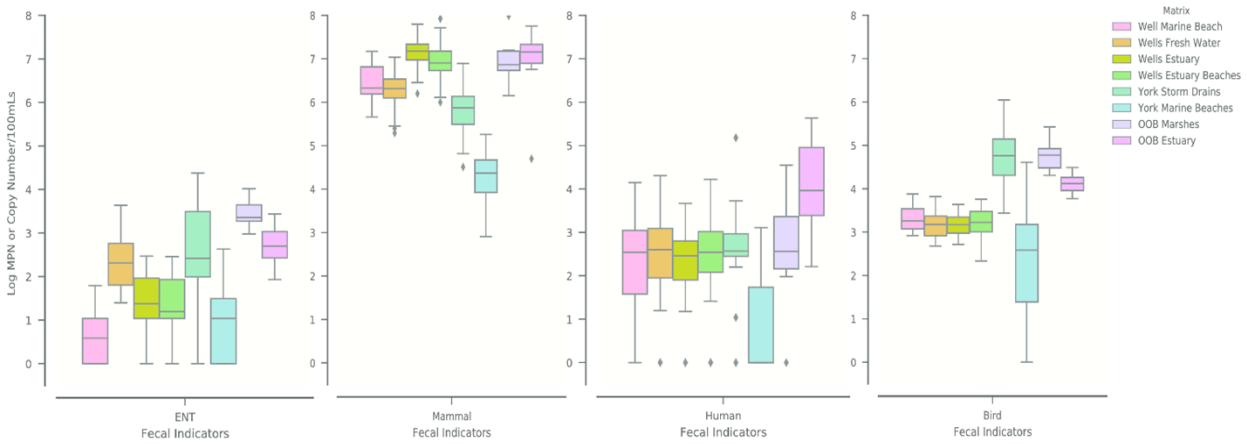


Figure 5. Regional Enterococci and Microbial Source Tracking Marker Concentrations. Colors correspond to sample water type and study locations are designated by the legend.

These regional comparisons enable developing a regional context for comparing MST marker concentrations to help determine their significance (Figure 5). For example, when bird fecal marker concentrations from Wells are compared to those from Old Orchard Beach (OOB), levels are significantly higher for OOB (Figure 5). This then provides a regional context for comparing bird fecal contamination levels between Wells, OOB, and any other area in the region, and suggests that fecal contamination from birds might represent more of a public health risk at OOB.

Overall, the use of MST to aid management decisions can be a powerful tool. Application of MST methods is ideally suited for areas that have shown chronically high levels of fecal contamination via fecal indicator bacteria concentrations, as application of the method without previous data to provide a context for MST studies might result in confusing and un-useful data. Identification of predominant and significant fecal sources is very important for management and reduction of TDMLs to recreational waters. However, an additional power in these MST methods is the relative quantification of markers. This allows for specific fecal sources to be tracked to a more refined geographic location by a traditional targeted sampling approach, only using MST marker instead of bacterial indicator detection. In the future, better-defined public health significance will be associated to each fecal marker, making relative concentrations more meaningful. In essence, MST can help managers make more well informed decisions on how to address water quality problems, and can help towns focus resources and remediation efforts.

1.5 Study Site Background and Research Questions.

In 2013 the National Research Defense Council conducted their most recent annual evaluation of the water quality at coastal beaches across the United States. Their findings resulted in Maine being ranked being ranked 27th (out of 30 states), with 11% of samples exceeding the regulatory limit (108). Although the State ranked poorly, Maine has a history of being proactive in addressing coastal recreational water quality issues. The Maine Healthy Beaches Program, which was established as part of an EPA funded program in cooperation with the Maine Department of Environmental Protection, has been working to address water quality issues since 2003. Their two-part mission is to monitor coastal recreational water quality across the State, and to educate and inform the public of relevant water quality issues. Through their work and previous work

done by the Jones lab, three study areas were chosen for investigation due to their histories of elevated level of fecal contamination, as measured by enterococci concentrations.

1.5.1 York, ME

The Town of York is a popular coastal beach area located in Southern Maine. The area sees a dramatic increase in population during the summer months, mostly because of its two most popular beaches, Long Sands and Short Sands. Given their locations and surrounding infrastructure, these areas of York contain storm drains that discharge directly onto its beaches. These storm drains contain untreated storm water, and are under regulations from the EPA's Municipal Separate Storm Sewer Systems (MS4) Program. These regulations seek to ensure untreated runoff (storm-water) to public recreational waters meets water quality standards, and if not a plan for remediation needs to be constructed. In 2014, the Town of York contacted the Jones lab to conduct intensive sampling in-order to develop a predictive model for water quality. As mentioned in Sections 1.4.1 & 1.4.2, findings identified storm drains as a significant source of enterococci, and that >1.5 inches of rain in 48 hours is likely to result in elevated concentrations of enterococci in the beach water (100). This then led to storm drains and their receiving beach waters to be the focus of an intensive MST study in 2015.

1.5.2 Wells, ME

The Town of Wells, ME is another coastal beach town in Southern Maine, however relative to York and Old Orchard beaches its summer increase in population is much lower. Within the town there are two marine beaches (Wells Beach & Drakes Island), as well as a large estuary that also contains public beaches. In the early 2000s, the Jones lab conducted MST studies in the

surrounding Webhannet watershed using ribotyping to identify sources of fecal contamination. Findings showed that freshwater inputs contained geometric mean concentrations of *E. coli* that exceeded >100 CFU/100 ml, and that human and wildlife were significant sources of fecal contamination (109). In 2014, the Maine Healthy Beaches reported that beaches within Wells Harbor had a >20% exceedance rate for enterococci concentrations related to the state standard, which they suggested as being influenced from freshwater inputs (101). This historic information led to Wells being a target of a 2016 MST study that sought to identify potential relationships between enterococci and detected source-specific MST markers in a new and comprehensive fashion.

1.5.3 Old Orchard Beach, ME

Old Orchard Beach can be considered the State's most popular summertime beach, with its economy almost entirely dependent on summer tourism. Work conducted by Maine Health Beaches and local towns from 2012–2014 identified the Goosefare Brook watershed to be significantly impaired with fecal contamination, based on enterococci concentrations (101). Specifically, the mouth of the Goosefare Brook and the upstream marsh area were considered the most problematic areas. Technical help from the EPA suggested that human fecal sources could be present, but no intensive MST studies had been conducted in the area. The culmination of these factors led to a partnership with the Maine Healthy Beaches Program in 2016, to conduct an intensive MST study in the contaminated Goosefare Brook tidal river and marsh area.

1.5.3 Research Questions

The work presented in this thesis is very applied. All field research projects were in areas that

had historically elevated levels of fecal contamination measured only by enterococci concentrations. The main applied research objective was to use new MST tools to identify predominant fecal sources present in problematic areas to inform management action. Secondary objectives were focused on identifying seasonal and spatial patterns of fecal contamination in each study area. Together, these objectives were designed to address practical needs that could inform better management strategies and remediation efforts. However, the focus of this thesis is also centered on more basic research questions that were complementary to the applied research objectives. The following are brief summaries and justifications for two research questions presented in this thesis.

As mentioned in Section 1.3, enterococci can persist and even grow in the environment. The role of environmental enterococci reservoirs has been discussed as a significant non-point source that can influence recreational downstream water quality (47, 110), however their roles have been investigated in a relatively narrow, non-integrative fashion thus far in the field. For example, studies that have investigated environmental sources of enterococci have either identified that a reservoir exists (54), used mesocosms to show enterococci persist longer in the presence of particles (53), or have determined resuspension dynamics in stream beds (62). To date, no comprehensive MST study has been conducted that analyzes how environmental source input, environmental conditions, and fecal source input could relate to enterococci concentrations in coastal recreational waters (Figure 6).

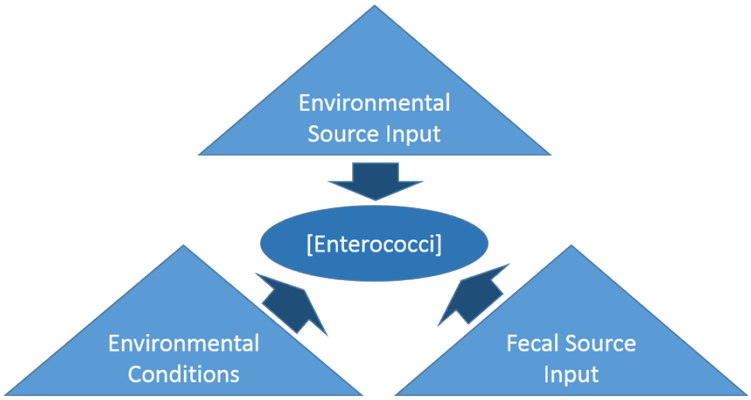


Figure 6. Potential Sources and Influences of Enterococci Concentrations in Recreational Waters.

Thus, the first research question is, "*Does detection of enterococci in a coastal beach and surrounding watershed relate to fecal source input, potentially from a variety of sources, or are enterococci detection and concentrations more related to environmental source inputs and conditions?"*.

Fecal contamination of recreational waters represents a public health risk. The original studies that evaluated public health outcomes in relation to enterococci concentrations found a relationship between the two when there was a prominent human fecal source contaminating the water (1–4). However, many beaches across the United States are not impaired predominately by point sources of human pollution, and instead non-point sources are significant contributors to beach fecal contamination loads (6, 10). Within the literature there is a significant gap in research that has evaluated recreational beaches that are impacted by multiple fecal sources. Thus, the second research question sought to investigate this by specifically asking, *"What is the relationship between different fecal sources identified in the environment and potential bacterial pathogens present? Do specific fecal sources relate to differential levels of potential*

pathogens?”.

Together, the research that addressed both sets of research questions sought to add valuable insight to the field of microbial source tracking. The major importance of this work lies in the fact that encompassing approaches and application of new technologies were used to answer research questions that are lightly covered in the field. The two research chapters presented here will ultimately be written for publication, so their formats follow those conventions.

Chapter 2: Influences of Fecal Sources, Environmental Sources and Environmental Conditions on Enterococci Concentrations in Coastal Water

Abstract:

Fecal pollution at coastal beaches in the Northeast, USA presents public health and economic loss concerns that require management efforts to address. Water quality is routinely determined using fecal indicator bacteria, specifically enterococci for marine recreational water. However,

enterococci concentrations can be influenced by a variety of variables including fecal deposition, environmental conditions, and environmental reservoir sources (i.e. sediments and soils), skewing their public health significance. In this study, we sought to delineate the influences of the above factors on enterococci concentrations in southern Maine coastal recreational waters. Weekly water samples and water quality measurements were conducted at freshwater, estuarine, and marine beach sites from June through September. Samples were analyzed for total and particle associated enterococci concentrations and molecular microbial source tracking (MST) markers for mammal, human, dog, bird, and ruminant fecal sources. Water samples and soil, sediment, and marine sediment samples were subjected to 16S rRNA sequencing and SourceTracker analysis to determine relative influences from these environmental reservoirs on water sample microbial communities. The study showed elevated levels of enterococci and particle-associated enterococci in freshwater, with total suspended soils relatively similar across all sites. Freshwater was a major conduit of bacterial transport, but the microbial community was also significantly influenced by sediment. Mammal fecal contamination was significantly elevated in the estuary, with human and bird fecal concentrations similar between sites. Out of all measured variables, particle-associated enterococci and mammal fecal concentrations had the most significant positive relationship with total enterococci concentrations within the estuary/marine beach and freshwater environments. Overall, elevated enterococci levels are reflective of a combination of increased fecal source input, environmental sources, and environmental conditions, which can skew their public health significance.

1. Introduction:

Fecal contamination of coastal recreational waters is a significant public health concern, as fecal material, often from nonpoint sources, can harbor an array of different pathogens. The US EPA has established regulations based on enterococci bacteria as the indicator of fecal-borne pollution to help manage water quality at estuarine and marine beaches (13). These organisms correlated well with predicted public health outcomes in several epidemiological studies that served as the basis for their adoption as the regulatory water quality indicator (4–7). The presence of human feces can present an elevated public health risk in recreational waters compared to non-human sources due to the lack of an "inter-species barrier" for diseases and the higher density of human pathogens that humans can carry (8–10). Although humans represent the greatest public health risk, other fecal sources that contain enterococci and possibly human pathogens can be chronic or intermittent sources of both, making beach water quality making management and remediation efforts more complex.

The need to differentiate fecal sources in recreational waters led to the emergence of microbial source tracking (MST) methods, most notability are the PCR-based assays that target the 16S rRNA in *Bacteroides* spp. (66, 75). There are a wide range of species-specific genetic markers that have been designed that range from human fecal source to various domestic and wildlife fecal sources. These assays have been in use for well over a decade and are supported by numerous and rigorous laboratory evaluations and field applications (70, 77, 81, 85, 111–113). Initial field studies investigated the relationship between MST markers and FIB concentrations in recreational waters to better elucidate potential sources of fecal pollution. Some studies have found strong relationships between the MST markers and enterococci (111, 114) while other

studies have found either weak or no relationships (115–117), all of which are discussed in a review by Harwood et al. (25). The main factor affecting the relationship between enterococci and the relative strength of different sources of fecal contamination is that enterococci can persist and grow in the environment, which can significantly influence their concentrations in recreational water (47).

Due to the promiscuous nature of enterococci in natural ecosystems, recent studies have been conducted to not only elucidate environmental parameters controlling their growth, but also to identify naturalized niches that can act as reservoirs for enterococci and the associated influence on water quality measurements. Specifically, enterococci have been shown to persist in fresh water sediments (48–50) and marine sediments (48, 51), and in some cases their relative concentrations in sediments are several orders of magnitude higher than the overlying water (48, 60, 61, 118). In addition, enterococci persist in soils affected by anthropogenic activities (52) as well as more natural soil environments (53–55). Both types of soil can act as significant reservoirs of enterococci that can confound concentrations observed in recreational waters. Evaluating the influence of sediment and or soil on water quality has, in some studies, been conducted by measuring total suspended solids as a surrogate for sediment-associated enterococci (51, 62, 119), however this approach is non-specific in that it does not indicate the specific type of source(s) for measured suspended solids. With the advent of next generation sequencing, sources of bacteria (i.e. sediments or soils) can be fingerprinted via 16S rRNA sequencing and programs like SourceTracker can then determine relative fractions of source-specific 16S fingerprints within a water sample (89).

This study examined the coastal and estuarine beaches of Wells, ME where there has been historically elevated enterococci levels, as reported by the Maine Healthy Beaches Program (101). Prior to this study, no recent MST studies had been conducted in this area, however, based on a ribotyping-based MST study (109) and other indicator tracking work, the two major freshwater inputs, the Webhannet River and Depot Brook were considered to be the major influences on water quality related to an array of fecal contamination sources. To investigate potential sources of enterococci we measured three major categories of variables (fecal source input, environmental conditions, and environmental sources) and then used a partial least squares regression model approach to determine the most significant influences on the enterococci concentrations in water samples.

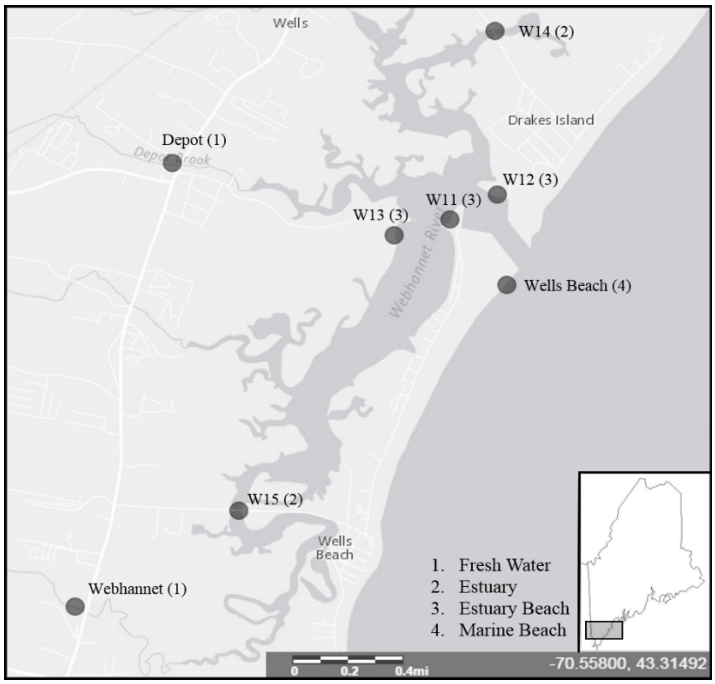


Figure 1: Wells Maine Study area and sampling sites. All water collection sites are marked with a dark grey circle. Sites that correspond to fresh water are indicated with a (1), estuary (2), estuary beach (3), and marine beach (4).

2. Materials and Methods:

2.1 Site description. This study was conducted in Wells, Maine, USA (Figure 1). Eight different sites were used to monitor water quality (n = 2 freshwater, n = 2 estuary, n = 3 estuary beaches, n = 1 marine beach) as well as twelve soil, twelve fresh-water sediment and four estuarine sediment sampling sites. Air temperature and rainfall data for the 48 h prior to sampling were obtained from Weather Underground (https://www.wunderground.com/cgibin/findweather/getForecast?query=Wells,%20ME) and characteristics of tides during sampling were obtained from www.meusharbors.com.

2.2 Water Sampling. Surface water samples were collected weekly from June to September 2017 (n = 117). Sampling started two hours before low tide to maximize the potential impacts of freshwater pollution sources, and all estuary and marine beach sites collected before the slack tide. Water samples were collected in autoclaved 1L Nalgene™ Wide-Mouth Lab Quality PPCO bottles (Thermo Fisher Scientific, Waltham, MA, USA), and environmental parameters were measured with a YSI Pro2030® Dissolved Oxygen, Conductivity, and Salinity Instrument (YSI Incorporated, Yellow Springs, Ohio, USA). A field replicate was collected at a different site for each sampling event.

2.2 Soil, Sediment, and Marine Sediment Collection. Environmental sources were collected twice throughout the sampling season to build source libraries that were "finger-printed" with 16S sequencing and SourceTracker analysis. Soil and sediment samples were collected upstream of both freshwater sites (6 Webhannet, 6 Depot). Soil samples were collected at the crest of the

stream embankment, where a 10 x 10 cm a plastic square template was placed down and all soil (O-horizon) within the template at a 2 cm depth was collected. Samples were sieved to remove any loose-leaf litter and roots to only sample smaller soil particles and their microbes. Underlying stream sediments were collected using a Van Veen sediment sampler from depositional sites chosen based on the presence of fine grain sediments. One grab sample was collected for each site and then the top 2 cm of sediment was subsampled for analysis. Sediments were sieved to remove coarse grain and gravel size particles. Estuarine sediments were collected during low tide when intertidal sediments were exposed using the Van Veen sampler, and the top 2 cm were again collected for analysis.

2.3 Enterococci and Total Suspended Solids Quantification. Total and particle-associated enterococci were enumerated using the EPA Method 1600 membrane filtration protocol (120) and particle-associated enterococci were determined via filtration through a 0.47 mm diameter 3.0 μm pore size polycarbonate filter (Millipore™, Darmstadt, Germany) as first reported by Crump et al. (121). The filters were rolled onto plates containing mEI agar and incubated at 41°C ± 0.5°C; representative colonies were counted in 24 ± 2 hours. Total suspended solids (TSS) were measured using EPA method 160-2, where 500 ml of the water sample was used to determine TSS concentrations (122).

2.4 DNA Extractions. DNA extraction from all matrices was performed with the PowerSoils® DNA Extraction Kits (MO BIO Laboratories, Carlsbad, CA, USA), with modifications to the manufacture's protocol needed to optimize the extraction from water sample filters. For water

samples, 500 ml collected water sample was filtered through 0.47 mm diameter 0.45 µm pore size polycarbonate filter (Millipore™, Darmstadt, Germany), which was stored in a sterile 2 ml cryotube at -80°C for at least 24 h. Prior to DNA extraction, frozen filters were crushed into small pieces with an ethanol sterilized razor blade, a practice commonly used to maximize DNA recovery (88, 123, 124). To minimize additional DNA loss during the extraction process solutions C2 and C3 (from manufacturer's protocol) were halved in volume and combined into a single step. DNA extraction from soil, freshwater sediment, and marine sediment were conducted per the manufacture's protocol.

2.5 Microbial Source Tracking (MST) PCR and qPCR Assays. MST PCR assays that target Mammals (Bac32; 37), Humans (HF183; 9), Gulls (Gull2; 37), Dogs (DF475; 31) and Ruminants (CF128; 9) were used to determine the presence of fecal sources in water samples. Positive control plasmids were made for each assay from an assay specific fresh fecal sample using the TOPO™ TA™ Cloning Kits (Invitrogen, Carlsbad, CA, USA), with a blue/white screen of constructs on kanamycin (50 µg/mL) selective TSA plates. White colonies were screened with their respective PCR assay, and positives were then grown in TSB and extracted with the PureLink® Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA). PCR assays were run on a T100™ Thermal Cycler (BioRad, Hercules, CA, USA) with the GoTaq® Green MasterMix (Promega, Madison, WI, USA). Cycling conditions and amplification protocols for each assay targeted the different source specific markers and followed protocols delineated by different studies: Bac32 and HF183 (125), CF128 (126), DF475 (127), and Gull2 (80). Quantitative PCR assays were also run to determine fecal source strength for Mammals (AllBac; 104), Humans (HF183; 29), and Birds (GFD; 40). All qPCR assays were run on a Mx3000P cycler (Agilent

Technologies, Santa Clara, CA, USA), TaqMan assays used the PerfecCTa® FastMix® II (QuantaBio, Beverly, MA, USA) master mix and the SYBR green assay used the FastSYBR™ Green Master Mix (Applied Biosystems, Foster City, CA, USA). A standard curve ranging from 10^6-10^1 copies was also run for each experimental run. The Ct values, amplification efficiency, slope, and R^2 values for each standard curve were compared to previously run standard curves, to ensure satisfactory performance before being used to calculate copies numbers for that run. Each environmental sample was diluted 1:10 and run in triplicate and the reaction volume (25 µl) contained a final concentration of 0.2 mg/ml BSA. Amplification/cycling conditions were preformed per published protocols for AllBac (129), HF183 (130), and GFD (85). TaqMan assays were run with an internal amplification control (131) with a down-shift of 1 cycle considered inhibition. Samples spiked with a plasmid containing 10^4 copies of GFD amplicon were used as inhibition controls for the SYBR assay, with a recovery of less than 10^4 copies (100%) considered inhibition. For a list of primers, probes, and standard curve performance, see Supplementary Material 1.

2.6 16S Library Preparation. The V4 region of the 16S rRNA gene, using the 515F-806R primer-barcode pairs, was used for amplicon sequencing (132). The Earth Microbiome Project protocol was used for amplification and pooling of samples, with minor modifications (133). The Qubit® dsDNA HS assay was used to quantify sample concentrations, and 500 ng of DNA was pooled per sample. The pool was then run on a 1.2 % low-melt agarose gel to separate primer-dimers from acceptable product, and bands between 300-350 bps were cut and extracted as described above. The final DNA sample was then run on the Agilent Technologies 2200 TapeStation system (Santa Clara, CA, USA) to determine final size, quality, and purity of

sample. Each library was sent to the Hubbard Center for Genome Studies at the University of New Hampshire to be sequenced (2 x 250 bp) on the Illumina HiSeq 2500 (San Diego, CA, USA).

2.6 Quality Filtering and OTU Picking. QIIME 1.9.1 was used to perform all major quality filtering, and OTU picking (Caporaso et al., 2010). Forward and reversed reads were quality trimmed (μ P25) and removed of Illumina adapters via Trimmomatic (135). Any reads that were less than 200 bps were discarded, and reads were merged with the QIIME joined_paired_ends.py, using a minimum overlap of 10 bps and a maximum percent difference of 10%. Paired-end data were analyzed using the QIIME open-reference OTU picking strategy with UCLUST for *de novo* picking and the Greengenes 13_8 database (136) for taxonomic assignment. Data for all samples are available on the Short Read Archive (http://www.ncbi.nlm.nih.gov/sra/) project reference (_____). Alternative OTU picking strategies were also tested to determine best workflow. For data output and performance of difference strategies refer to Supplementary Material 2.

2.7 SourceTracker Analysis. Samples from 4 source types (fresh water, soil, sediment, and marine sediment) and 4 sink types (fresh water, estuary water, estuary beach water, and marine beach water) were analyzed by the open-source software SourceTracker v1.0 (Knights et al., 2013). Default parameters were used (rarefaction depth 1000, burn-in 100, restart 10, alpha (0.001) and beta (0.01) dirichlet hyperparameters). A 'leave one out' cross validation was preformed to assess the general performance of the model and source samples were iteratively

assigned as sinks to assess how well a known sink would be assigned (i.e. source = soil and sink = soil). The percent assignments from SourceTracker are the result of the Gibbs Sampler assigning OTUs from an unknown sample to sources in a random and iterative fashion, and then calculating likelihood of that OTU originating from said source. The final output can be interpreted as the percent (or likelihood) of OTUs present in an unknown sample originating from the sources used in the analysis

2.9 Partial Least Squares Regression Model. A partial least squares regression (PLSR) model was used to distill out the most important and significant variables affecting enterococci concentrations (137). Two models were created, one for the estuary, estuary beach, and marine beach sites, and one for the freshwater sites. Particle-associated enterococci, environment variables (water temperature, air temperature, dissolved oxygen, salinity, height of previous high tide, rainfall in previous 48 h), fecal source strength (mammal, human, and bird), and percent of environmental source (fresh water, soil, sediment, and marine sediment) were all used as explanatory variables for the non-freshwater model. The same parameters, except height of previous high tide and percent of freshwater source, were used for the freshwater model. All data except the percent assignments from SourceTracker were log (x+1) transformed before performing the analysis. A KFold cross validation (K=7) with the NIPALS method was used to determine optimal factors and variable importance (VIP > 0.8) for each model. Models were then re-run with only explanatory variables that were determined to be significant. To see model validation and diagnostic plots, refer to Supplementary Material 3.

2.10 Routine Statistical Analysis and Data Visualizations. All routine statistical analyses were performed in R v3.4.0, Python 3.6.1, or JMP Pro13, while multivariate analyses were performed with PC-ORD v6. Graphing was performed in IPython notebook with matplotlib, seaborn, pandas, and numpy packages. All pairwise comparisons were done using the Kruskal-Wallis nonparametric method, with Dunn's nonparametric multiple comparisons run *post hoc* using a Bonferroni correction.

3 Results:

3.1 Total and Particle-Associated Enterococci Concentrations and Total Suspended Solids in Water. During this study, total enterococci concentrations were highest in freshwater sites, with concentrations significantly decreasing from there to the estuary and then the marine beach areas (Figure 2). In addition to measuring enterococci concentrations in water samples, particle-associated enterococci and suspended solid concentrations were measured to better understand the potential mode of transport of these bacteria within this coastal watershed. Throughout the study period (June-September 2016), levels of total and particle-associated enterococci varied by site. Concentrations were lowest at the marine beach (Wells Beach) compared to others sites, with levels significantly higher in estuary (W11, W12, W13) and freshwater sites (Depot & Webhannet; Figure 2).

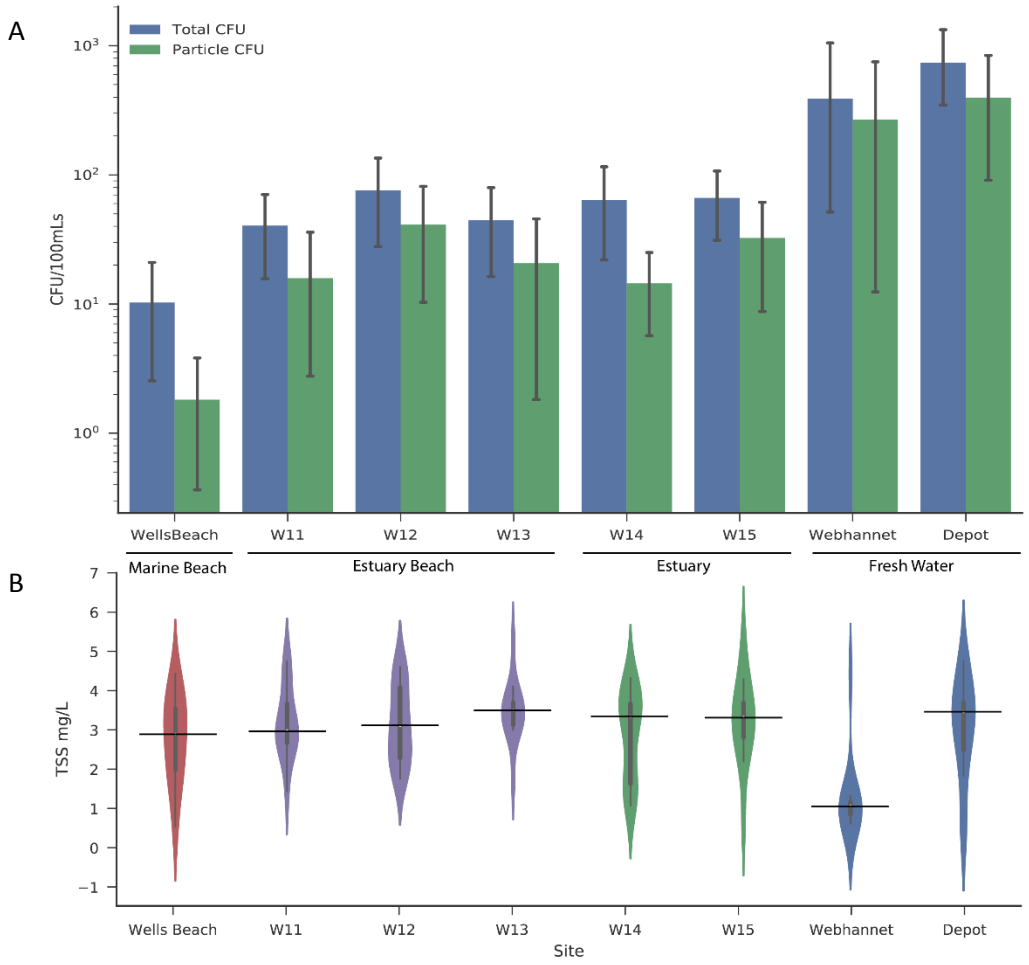


Figure 2: Geometric Mean Concentrations of Total and Particle Associated Enterococci and Average Total Suspended Solids Concentrations at the Eight Study Sites. (A) Total enterococci concentrations are represented with the blue bar, and particle associated enterococci concentrations correspond to the green bar. Error bars are derived from variation from each site across the entire study. (B) Violin plots were used to represent TSS concentrations, and the color corresponds to the type of site including marine beach (red), estuary beach (purple), estuary (green), or fresh water (blue). Horizontal lines go through the median of each violin plot.

Both total and particle-associated enterococci geometric mean concentrations were statistically similar at the estuary beach (W11, W12, W13) and estuary (W14, W15) sites. Freshwater sites (Webhannet and Depot) however, had statistically higher enterococci concentrations than other sites (Figure 2; $p < 0.05$). The ratio of total to particle-associated enterococci varied throughout

the season, with an average of 36.3% (SD ± 30) across all sites. Sites within the estuary beach showed the highest ratio (41%, SD ± 32), however there were no significant differences observed between sites or types of sites. Average TSS concentrations were similar for most sites site, with an overall average of 2.9 mg TSS/L (SD± 1.2), however, the Webhannet freshwater site had significantly lower average TSS concentrations (1.23 mg/L ± 1.0SD, $p < 0.05$) (Figure 2), despite, as previously mentioned, having higher enterococci concentrations. The relationship between particle-associated enterococci and TSS was not significant ($r^2 = 0.0011$). Overall, this study showed enterococci concentrations differed most significantly by site and were ubiquitously associated with particles, which was independent of suspended solids concentrations.

3.2 Presence of Fecal Sources in Fresh, Estuarine, and Marine Waters. The impact of fecal pollution in this study area was determined using both PCR and quantitative PCR MST assays to identify and quantify predominant sources of fecal contamination present in the water. The mammal fecal marker (Bac32) was always detected via PCR at all sites and through the whole study period. (Supplementary Material 1E). The human fecal marker (HF183) was detected in 51% of all water samples, with the highest detection rate in fresh water (56%) and the lowest detection rate in marine beach water (46%). Differences in the percent detection of the gull fecal marker (Gull2) were most pronounced between freshwater (10%) and all other sites (>77%). The dog fecal marker (DF475) detection rate was highest in the estuary beach water (10/44 = 23%), however 8 of the 10 positive samples were detected in July (8/13 = 61%). For all other sites, an increase in the detection of dog fecal marker also occurred during July, with 44% (16/36)

detection, compared to 0% for August and September and <1% for June. Thus, most of the dog contamination at all sites was associated with dog-related conditions during July.

Concentrations of Mammal, Human, and Bird Fecal Sources. We used qPCR to provide relative quantitative measures of mammal, human and bird fecal contamination levels. Water at the estuary and estuary beach sites contained significantly higher levels of mammal (AllBac) fecal marker copies, with an average of 1.54×10^7 compared to 2.62×10^6 in freshwater and 3.9×10^6 copies/100 ml in marine beach ($p < 0.05$). Average concentrations of human (HF183) and bird (GFD) fecal markers were not statistically different between sites, however, concentrations of the human marker in individual samples varied from 0 - 2.04×10^4 copies/100 ml (Figure 3), while bird fecal marker concentrations were relatively stable across all sites.

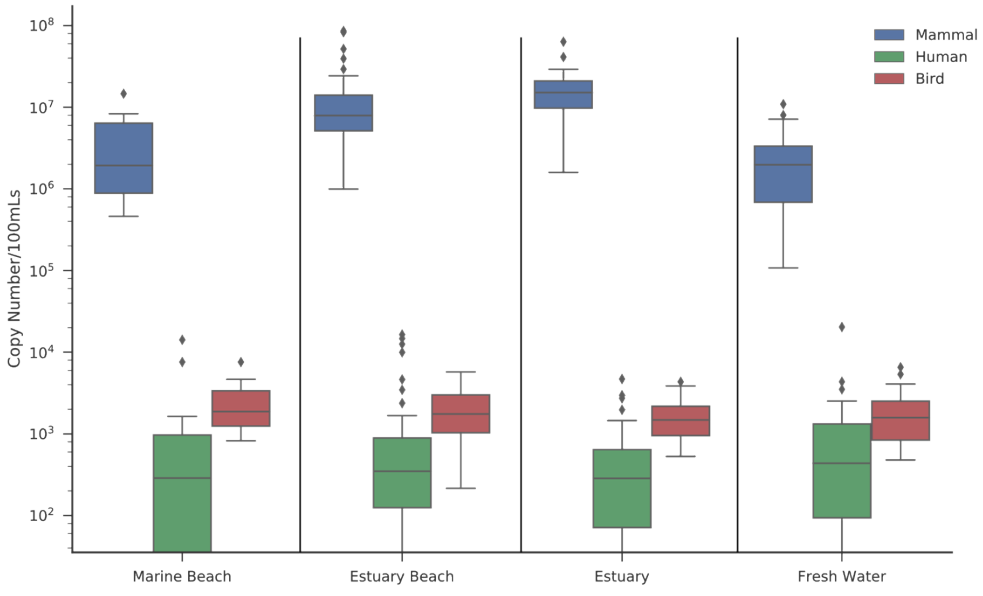


Figure 3: Relative Levels of Mammal, Human, and Bird Fecal Source at the Different Types of Study Sites. Box plots represent levels of microbial source tracking markers at marine beach (Wells Beach), estuary beach (W11, W12, W13), estuary (W14 & W15), and fresh water (Webhannet & Depot). Outlier data are represented with a black diamond.

No significant temporal trends were observed for any of the quantitative fecal marker levels. Compared with presence/absence detection of fecal sources, quantitative measurements also did not show strong spatial patterns, except mammal marker levels showed significant increases at estuary and estuary beach sites compared to marine and freshwater sites.

3.3 Differences Between Water, Soil, and Sediment Bacterial Community Compositions. 16S amplicon sequencing was used to characterize the microbial community present in water and other sample matrices (soil, sediment, and marine sediment), which was the nexus for ensuing SourceTracker analysis. A total of 3,276,196 reads and 7,706 unique OTUs were obtained from the 177 samples of fresh, estuary, estuary beach and marine beach water and soil, sediment, and marine sediment. The number of OTUs assigned and the Shannon diversity index were significantly higher for soil, sediment, and marine sediment when compared to water samples (Supplementary Material 2, $p < 0.05$). Most taxa in the estuary and marine beach samples were identified as Flavobacteriia, Alphaproteobacteria and Gammaproteobacteria classes, which together accounted for 84% of the total assigned taxa. Cyanobacteria accounted for 34% of the taxa in marine sediment, and Betaproteobacteria was one of the top three most abundant taxa in fresh water, soil and sediment (Figure 4).

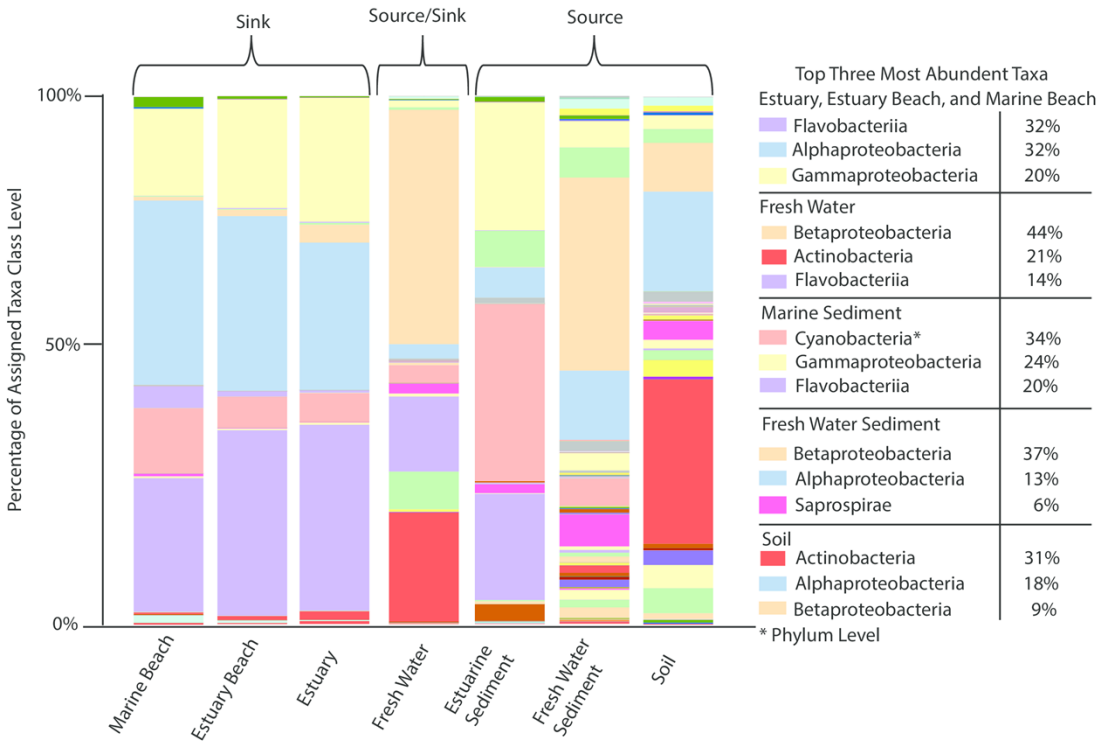


Figure 4: 16S Taxa Profiles and the Top Three Most Abundant Bacterial Classes in All Source and Sink Samples. Stacked bar plots represent percentages of the class level composition of the microbial communities. Source corresponds to environmental sources that were finger-printed with the SourceTracker program, and then used to determine their presence within water (sink) samples. The table represents the top three classes for each group of samples and * corresponds to phylum level. For a complete list of all taxa assignments refer to Supplementary material 4.

A Non-Metric Multi-Dimensional Scaling (NMDS) ordination was used to determine if the bacterial communities from water and other matrices (soil and sediments) differed based on their taxonomic composition. Bacterial communities from the marine beach and estuary (All Estuary) waters were similar, but were statistically different from fresh water (Figure 5, $p < 0.05$). The bacterial communities associated with soil, sediment and marine sediment were all distinct when compared to each other and water samples, indicating unique groups of OTUs (Figure 5, $p < 0.05$).

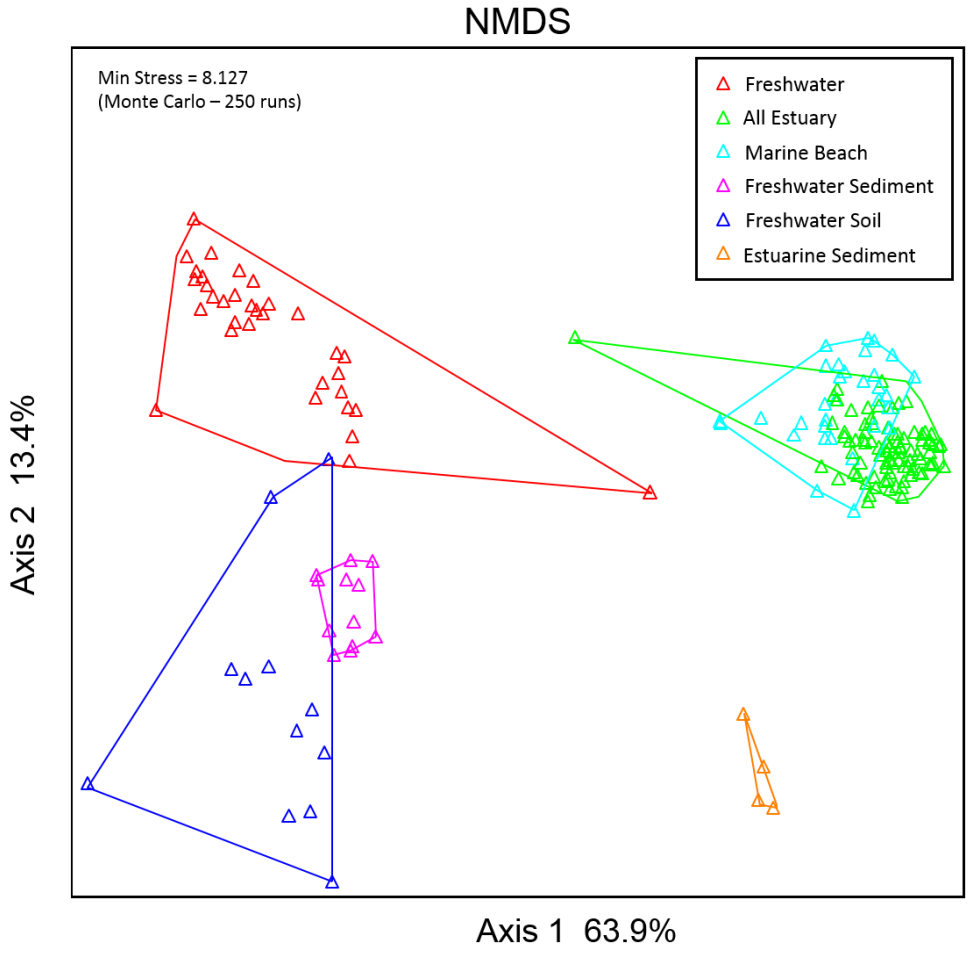


Figure 5. Differences Between Microbial Communities from Different Source Materials. Samples are color-coded based on sample matrix (i.e. soil, fresh water, etc.). Percent of variation explained are displayed on the x and y axis and the minimum stress of the ordination is shown in the top left corner.

Samples taken from different areas within the watershed (soil, estuarine water, freshwater, etc.) contained unique bacterial compositions, allowing for downstream analysis with the

SourceTracker software to discern relative contributions of these different communities to the make-up of microbial communities in the different types of water samples.

3.4 Environmental Source Contribution to Water Samples. The fraction of freshwater, sediment, soil, estuarine sediment, and marine beach water source bacterial communities within estuary and estuary beaches water samples were calculated using the Bayesian mixing model SourceTracker. Results indicated that marine beach water was the dominant source of bacteria in the estuary and estuary beach (Table 1).

Environmental Microbial Community Sources Including Marine Beach Source

Water Sample Type	Marine Beach	Freshwater	Estuarine Sediment	Sediment	Soil
Estuary Beach	97%	<0.01%	0.4%	<0.01%	<0.01%
Estuary	94%	2.9%	0.2%	0.02%	<0.01%
Freshwater	<0.01%	N/A	<0.01%	74%	2.6%

Environmental Microbial Community Sources Excluding Marine Beach Source

Water Sample Type	Marine Beach	Freshwater	Estuarine Sediment	Sediment	Soil
Estuary Beach	N/A	66%	12%	<0.01%	<0.01%
Estuary	N/A	73%	7.6%	<0.01%	<0.01%
Marine Beach	N/A	35%	19%	<0.01%	<0.01%
Freshwater	N/A	N/A	0%	74%	2.6%

Table 1. **The relative contribution of different sources to the microbial communities in estuarine and marine water**. SourceTracker was run with two different configurations, one where Marine Beach water was included as a potential source (top) and a second run where Marine Beach water was excluded as a potential source (bottom).

However, given that likely fecal sources are coming from the watershed, we excluded marine beach water as a potential source and included it as a sink then re-analyzed the data. These second results showed that fresh water taxa had a high probability of being a significant fraction

of estuary (73%), estuary beach (66%) and marine beach (35%) water communities, with a significantly higher percentage for the estuary locations compared to the marine beach (Table 1, $p < 0.05$), which is more influenced by ocean microbial taxa.

The percent of unidentifiable taxa assignments showed an inverse relationship compared to percent of identified and assigned freshwater taxa, as unidentifiable assigned taxa in the marine beach was significantly higher (46%; $p < 0.05$), which is not surprising given that marine beach water community would likely be most influenced by non-terrestrial sources. Estuarine sediment was the highest likely identified source in the water from the marine beach site (19%), and it was significantly higher than percentages calculated for the estuary sites ($p < 0.05$). Freshwater sample analysis showed a high probability of OTUs originating from underlying sediment (74%) and much lower probability of OTUs originating from soil (2.6%). Overall results showed that freshwater source-related taxa were a pervasive source throughout the estuary and marine beach, with sediment source-related taxa most highly abundant within the fresh water.

3.5 Relationships Between Environmental Conditions, Fecal Source Concentrations, Environmental Sources and Enterococci Concentrations. Two PLSR models were created to determine relationships between enterococci and fecal source concentrations, environmental sources, and environmental conditions (outlined in Methods 2.9). The first 'freshwater' PLSR model indicated particle-associated enterococci concentration, concentration of mammal fecal marker, TSS concentration, percent of sediment source, percent of unknown source, and salinity were important variables (VIP > 0.8) in resolving variation in enterococci concentrations (Table

1). A one-factor (single PLSR regression) model was deemed optimal (root mean PRESS = 0.735), and showed that all variables (except salinity) had positive associations with enterococci concentrations. Values for model performance ($R^2Y = 0.6$, $R^2X = 0.5$, and $Q^2 = 0.4$) indicated that the model fit the data moderately well ($R^2X \geq 0.5$) but had poor predictive capability of enterococci concentrations ($Q^2 < 0.5$; Supplementary Material 3). Out of all of the important variables, particle-associated enterococci (Particle ENT) concentrations showed the strongest relationship to total enterococci concentrations (Table 2).

Fresh Water PLSR 1	
X Variable	Loading
Particle ENT	0.501
qPCR Mammal	0.352
TSS	0.408
% Sediment	0.336
% Unknown	0.476
Salinity	-0.344
Total Y Variance	60.1%

Estuary, Estuary Beach & Marine Beach PLSR 1		PLSR 2	
X Variable	Loading	X Variable	Loading
Particle ENT	0.456	Particle ENT	0.420
qPCR Mammal	0.438	qPCR Mammal	-0.337
% Fresh Water	0.408	% Fresh Water	-0.418
% Unknown	-0.457	% Unknown	0.389
Water Temp (C)	0.302	Water Temp (C)	-0.123
Hightide (ft)	0.170	Hightide (ft)	0.456
% Estuarine Sediment	-0.294	% Estuarine Sediment	0.401
Total Y Variance	47.2%	Cumulative Y Variance	61.8%

Table 2. Most Significant Relationships/Contributions for All Factors to Enterococci Concentrations. Shown is the output from a partial least squares regression for a freshwater and estuary/marine model. All variables shown have significant relationships for each model (VIP > 0.8), and loadings are derived from re-running models with only variables deemed significant. Model loadings are specific weights on a multivariate regression axis, positive and negative loadings refer to positive or negative relationships to enterococci concentrations. Negative loadings in the model are designated with a – before the number.

The second PLSR model, a two-factor/two PLSR regressions model, was the best fit (root mean PRESS = 0.744) from the PLSR constructed for the estuary, estuary beach, and marine beach sites. The analysis identified particle-associated enterococci concentration, mammal fecal source concentration, percent of freshwater, unidentified and estuarine sediment sources, water temperature, and high tide height as significantly related to enterococci concentrations. Factor one showed that all variables were positively associated, except for the percent unidentified and marine sediment sources. The second factor showed mammal fecal sources, freshwater sources, and water temperatures were negatively related to enterococci concentrations, which was the opposite of their associations for factor one. The high tide height and marine sediment were positively related to enterococci concentrations for factor 2 of the PLSR (Table 2). Together both factors explained 61.8% in the variation observed in enterococci concentrations, and model performance ($R^2Y = 0.6$, $R^2X = 0.5$, and $Q^2 = 0.6$) indicated better predictive ability with a similar fit to the data compared to the freshwater model (Supplementary Table 3). Out of all the potential variables measured (19 total) across three categories (fecal source input, environmental source contribution, and environmental conditions), particle-associated enterococci and mammal fecal marker concentrations had the most significant relationships to enterococci concentrations. The relationships between other variables and enterococci concentrations were specific to freshwater and estuary/marine beach models, indicating ecosystem specific relationships. However, the joint relationship of particle-associated and mammal fecal marker across freshwater and estuary/marine environments indicate their overarching importance in determining enterococci concentrations.

4 Discussion:

Enterococci concentrations at the marine beach, estuary, and estuary beaches were, on average, lower than the 104 CFU/100 ml water quality standard for the State of Maine, suggesting these waters were often safe for recreational use. Previous monitoring by the Maine Healthy Beaches Program had shown the Wells Beach area was one of 7 beaches in Maine that had a greater than 20% exceedance rate and collectively with the other six beaches accounted for 30% of the total exceedances recorded program-wide, with suspicion that freshwater inputs were a significant source (101). Our findings confirmed that enterococci concentrations were statistically higher at both of the major freshwater tributaries to the estuary, where levels were on average above the 104 CFU/100 ml limit, suggesting the presence of significant sources of enterococci. However, the summer of 2016 was especially dry, with only one sample event coinciding with >1 inch of rain 48 h prior (1.73 in., 6/28/16) to the sampling time. This overall dry condition could have decreased fecal contamination transport (via freshwater discharge) from the watershed to the estuary and marine beach, thus lessening their impact on water quality, and suggests that because more average rainfall conditions would probably have caused more freshwater discharge, a higher loading of particle-associated enterococci would have probably impacted water quality.

Enterococci were significantly associated with suspended particles of >3 μm diameter. On average, 36% (SD ± 30) of the total enterococci concentrations were associated with particles, which suggests particles as a potentially important transport mechanism. Other studies conducted in estuary and storm waters have found similar fractions of particle associated enterococci, but they noted enterococci had a preference for a larger particle size of >30 μm (110, 138, 139). The large standard deviation for particle-associated enterococci could be attributed to the complex

nature of particle interactions (sedimentation rate, electrostatic, hydrophobic, and other surface-surface interactions) and hydrogeological dynamics (salinity-driven turbidity maximum) (110, 140). The mechanisms underlying enterococci-particle interactions may also be related to ionic strength in surface waters, as *Enterococcus faecalis* is negatively charged over a broad pH range (2-8 pH units) and in the presence of different ion concentrations (141). Results here showed that TSS and particle-associated enterococci had no linear relationship, indicating particle-associated enterococci were not dependent on the amount of suspended material and thus the association is likely due to other factors influencing cell-particle interactions.

Quantitative PCR assessment of several fecal sources are potentially useful measures to indicate relative significance of the different sources in a single sample and over time at sites of interest. PCR detection showed a chronic presence of mammalian fecal source (100% of samples) with human fecal source detected in about half of all samples, so it is important to put those findings into context with qPCR analyses. Mayer et al. (44) showed that wastewater effluent contains about 10^8 copies/100 ml of the AllBac mammal fecal marker, Sowah et al. (30) showed that streams impacted by septic systems could contain 10^5 – 10^7 copies/100 ml depending on the season, and Bushon et al. (142) showed that under storm flow conditions in an urban watershed mammal marker copy numbers could exceed 10^8 copies/100 ml. Results for this study ranged from 10^5 to 8.6 x 10^7 copies/100 ml, which is within previously reported ranges and likely a concentration reflective of a non-urbanized watershed. The estuary and estuary beach area showed a statistically higher concentration of the mammal marker, however, there was no increase in the concentrations of the human associated fecal marker (HF183), which is probably a reflection of humans not being the primary source for the increased fecal contamination. Also,

the average concentration of the human marker was 1,500 copies/100 ml. Boehm et al. (107) showed that 4,200 copies/100 ml of HF183 is the cutoff for where there are over 30/1000 predicted GI illnesses for swimmers (107). Although human sources are the greatest public health concern (8, 9, 25, 143), other mammalian fecal sources are more influential in explaining the variation observed in this study. Interestingly gull fecal sources were detected in 77% or more of the samples in the estuary and marine beach area, however only 10% of the samples were positive within the fresh water (Supplementary Material 1), despite there being no decrease in the bird fecal marker concentration, suggesting the presence of different bird sources in these areas. Anecdotally, Canada geese were observed upstream of both the Webhannet and Depot freshwater sites periodically throughout the season, which could be a significant source of bird fecal contamination within the fresh water (84).

One of the unique findings of this study was to determine the relative contribution of different sources to the bacterial community in the estuarine water. The bacterial community in estuarine water was mostly (>90%) from marine beach water, which is not surprising for a well-flushed estuary like the study site, and because the study period was minimally influenced by rainfall and associated runoff of freshwater, we expected that the influence of freshwater sources would be low. In ensuing analyses, we chose not to include marine beach water as a potential source for a variety of reasons. First, the samples were collected during low tide before the ebb when the estuary water was draining and water was moving from the watershed towards the marine beach. Secondly, we had already shown that the OTU compositions for the marine beach and estuary samples were very similar, increasing the possibility of a type I error (false positive). Lastly, sources are most likely to come from the watershed and not from marine water, so excluding

marine beach water helps to enhance the determination watershed influences. Our second analysis (marine beach source excluded) showed that freshwater was a significant source of bacteria to the estuary (>65% assignment) compared to soil, sediment, and estuarine sediment. Thus implicating freshwater as a major conduit for bacterial transport, as well as the major source of enterococci to the estuary. Overall this finding highlights the importance of freshwater discharge, as a controlling factor, in transporting contamination from the watershed to the coast. The specific percent assignment of freshwater source could be an overfit, however the trend observed is a very likely scenario given the above rational discussed.

Analysis of environmental reservoirs of enterococci (soil, sediment, etc.) and their presence within water samples using SourceTracker revealed a variety of source contributions to freshwater, estuary and marine waters. To date there have been limited studies using SourceTracker to identify soil and sediment-associated taxa within water samples, and none of these studies have focused on a coastal watershed with the potential for freshwater, estuarine and marine sources. One study conducted in the Minnesota River identified up to 14% of sediment and 1.4% of soil sources of the taxa within the river water (92). This study, however showed that the sediment source was much more abundant at freshwater sites (74%), indicating a much greater degree of mixing between the freshwater and sediment communities. The amount of sediment and soil sources within water samples may be related to a site specific characteristic like relief or soil texture, which has been shown with TSS fluxes on a global scale (144). Thus, the degree to which the underlying sediment community mixes with the overlaying water is likely site specific. Interestingly, even though freshwater contained a significant amount of sediment source taxa, no sediment source was observed at the estuary and marine beach sites

through SourceTracker analysis. This difference could indicate that rapid sedimentation happens during transit to and within the estuary and at the estuarine turbidity maximum zone (145). TSS concentrations and the ratio of particle-associated to total enterococci concentrations, however, showed no differences between freshwater and estuary/marine sites. This could be related to the separate and quite different hydrodynamics within these different water systems. The percent of sediment source in the freshwater samples observed here might also be an over-estimate/over fit from SourceTracker given the limited number of potential sources we used, but results consistently showed an elevated presence of sediment in all freshwater samples in this study. SourceTracker analysis also revealed that the freshwater source was significant (35% or more) in estuary and marine beach water samples, suggesting that fresh water is a significant conduit for microbial, and fecal contamination, transport from the watershed to the estuary and marine beach.

There has been growing use of partial least squares regression in the analysis of data from water quality studies, especially to construct predictive models for fecal indicator bacteria (146, 147). PLSR has been shown to significantly outperform other commonly used regression analyses, such as multiple linear regression and principle components regression (148). Results from the PLSR analysis in this study showed that particle-associated enterococci and concentrations of mammal fecal sources were the driving force behind variation in enterococci concentrations, as described by both of the PLSR models we constructed. Other factors were found to influence enterococci concentrations, however, these differed between the freshwater and estuary/marine beach models. For example, TSS concentration, the percent of both freshwater sediment and unknown sources positively influenced enterococci concentrations at freshwater sites. This

signifies that sediment is a likely source of enterococci that influences concentrations measured in the water. Positive influences from the unidentified source taxa indicates that there is either an alternative source (not measured in this study) within the watershed that also influences enterococci concentrations or that SourceTracker could simply not resolve all of the potential sources we used. This finding is not surprising given the vast number of potential sources of fecal pollution within a watershed and that fecal sources were not a part of the SourceTracker analysis. Results from the estuary and marine beach model returned a two-factor regression, with each factor essentially being the inverse of each other. Specifically, it highlighted freshwater being a major conduit for microbial transport to and through the estuary. Negative influences from the unknown source reaffirms this notion, along with positive influences from the previous high tide height. The second factor explained about 15% of the variation in enterococci concentrations, so it's importance must be weighed proportionately to factor one, which explained almost 50% of the variation. However, positive loadings from previous high tide height and percent of marine sediment indicate marine sediment could be a source of enterococci whose influence is dependent on tide height. The negative loadings from mammal fecal source may indicate that enterococci coming from the estuarine sediment are not from mammal fecal sources.

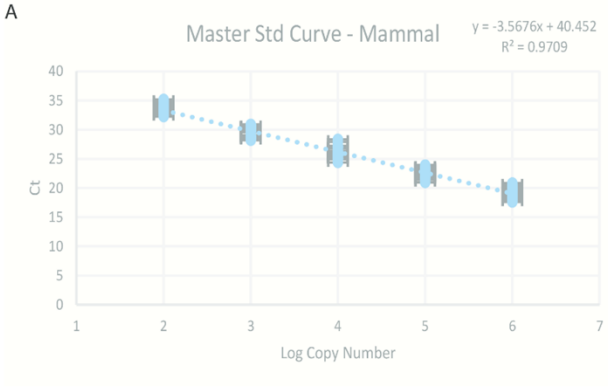
Overall, the results from this study indicated concentrations of enterococci in a coastal estuarine/marine beach environment are largely controlled by particle-associated enterococci and mammal fecal source input. Their influence is likely universal across freshwater and estuarine environments, however other ecosystem factors play a role as well. For freshwater parts of the coastal watershed, sediment could act as a significant enterococci reservoir that is constantly

being re-suspended within the water column. Freshwater itself could act as a major conduit for bacterial transport to an estuary and marine beach area where other environmental factors (water temperature and high tide height) can influence enterococci concentrations as well. These findings highlight the dynamic nature of enterococci in natural aquatic ecosystems outside of the mammalian fecal tract, and that concentrations within fresh water and estuary/marine beach water are influenced by variety of different factors.

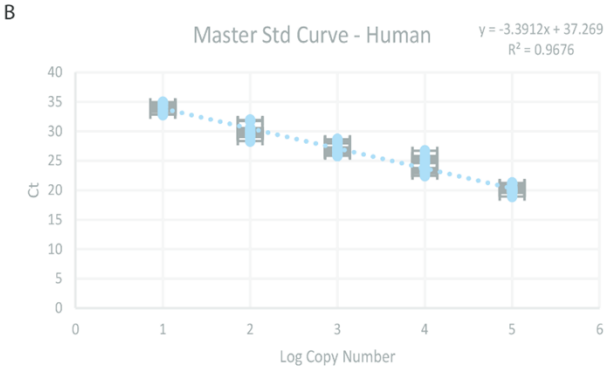
Supplementary Material 1: MST PCR Primer List, PCR Percent Detection, And qPCR Standard Curve Performance

Supplementary Material 1 - Master qPCR curves and MST assay primers and probes

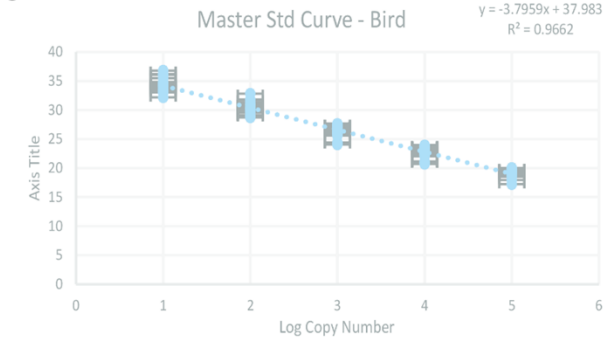
A



B



C



D

PCR Assays	Source	Taxa	Target/Size	Primers/Probes 5'-> 3'
Bac32	Mammal	*Bacteroides-Provotella*	*16S rRNA* (696bp)	Bac32: AACGCTAGCTACAGGCTT Bac708: CAATCGGAGTTCTTCGTG
HF183	Human	Human Cluster *Bacteroides-Provotella*	*16S rRNA* (541bp)	HF183: ATCATGAGTTCACATGTCCG Bac708: CAATCGGAGTTCTTCGTG
CF128	Ruminant	Ruminant Cluster *Bacteroides-Provotella*	*16S rRNA* (595bp)	CF128: CCAACYTCCCGWTACTC Bac708R: CAATCGGAGTTCTTCGTG
DF475F	Dog	Dog Cluster *Bacteroides-Provotella*	*16S rRNA* (251bp)	DF475: CGCTTGTATGTACCGGTACG Bac708: CAATCGGAGTTCTTCGTG
Gull2	Gulls	*Catellicoccus marimammalium*	*16S rRNA* (412bp)	Gull2F: TGCATCGACCTAAAGTTTTGAG Gull2R: GTCAAAGAGCGAGCAGTTACTA
qPCR Assays	**Source**	**Taxa**	**Target/Size**	**Primers/Probes 5'-> 3'**
AllBac			*16S rRNA* (108bp)	AllBac296f: GAGAGGAAGGTCCCCCAC AllBac412r: CGCTACTTGGCTGGTTCAG AllBac375Bhqr: (FAM)TGAAGGATGAAGGTTCTATGGATTGTAAACTT(BHQ-1)
TaqMan	Mammal	*Bacteroides-Provotella*		
HF183			*16S rRNA* (167bp)	HF183: ATCATGAGTTCACATGTCCG BDFRev: CGTAGGAGTTTGGACCGTGT
TaqMan	Human	Human Cluster *Bacteroides-Provotella*		BFDFAM: (FAM)CTGAGAGGAAGGTCCCCCACATTGGA(BHQ-1)
GFD SYBR	Avain	Unclassified *Helicobacter spp.*	*16S rRNA* (123bp)	F: TCGGCTGAGCACTCTAGGG R: GCGTCTCTTTGTACATCCCA

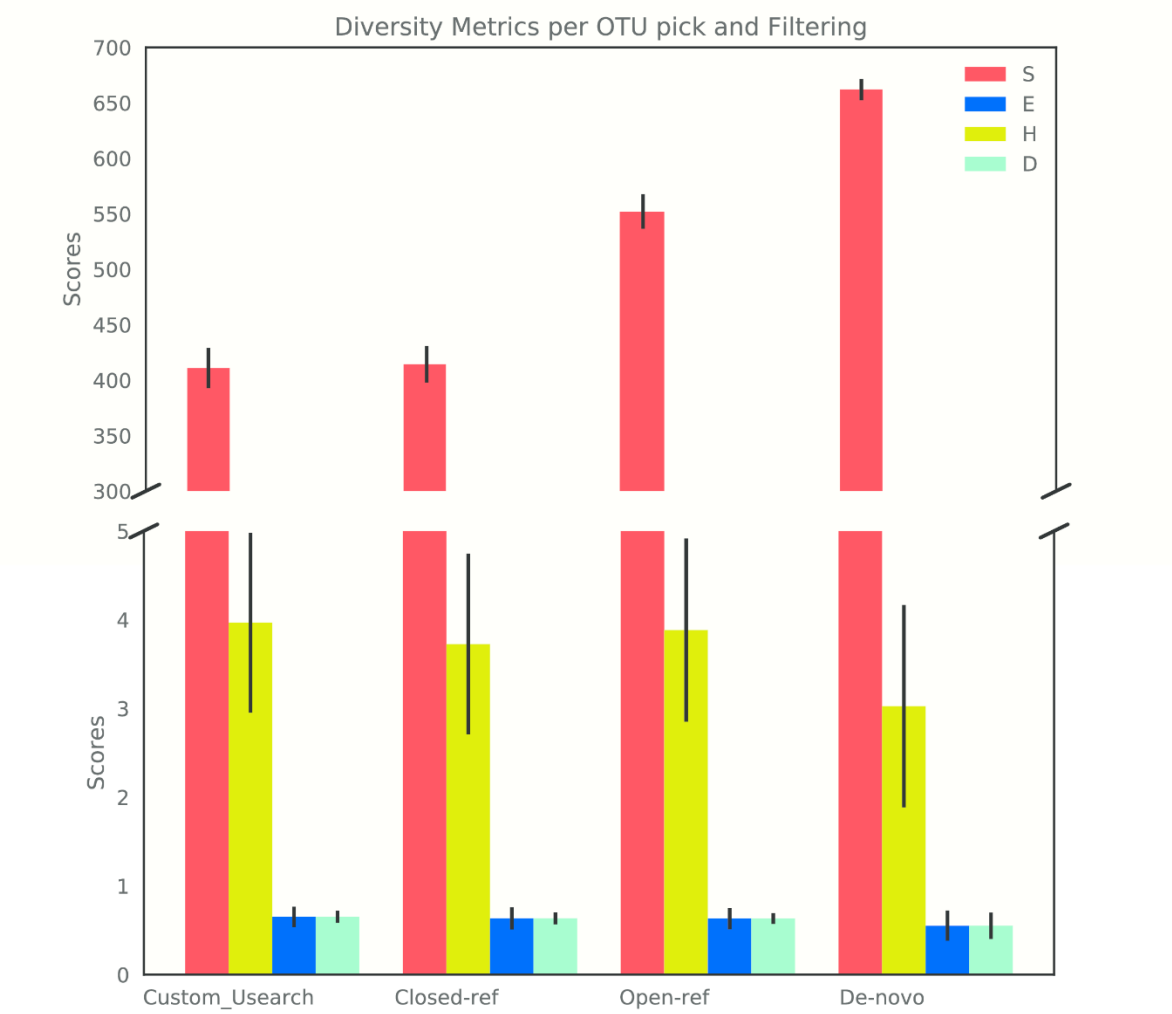
Supplementary Material 1 - MST PCR percent dections

E

MST PCR Percent Detection					
Fresh Water					
	Mammal	Human	Gull	Dog	Ruminant
n =	30	30	30	30	30
positives	30	17	3	3	0
percent	100.00	56.67	10.00	10.00	0.00
Estuary Beach					
	Mammal	Human	Gull	Dog	Ruminant
n =	44	44	44	44	44
positives	44	23	38	10	0
percent	100.00	52.27	86.36	22.73	0.00
Estuary					
	Mammal	Human	Gull	Dog	Ruminant
n =	30	30	30	30	30
positives	30	14	24	3	0
percent	100.00	46.67	80.00	10.00	0.00
Marine Beach					
	Mammal	Human	Gull	Dog	Ruminant
n =	13	13	13	13	13
positives	13	6	10	2	0
percent	100.00	46.15	76.92	15.38	0.00

Supplementary Material 2: Diversity Measurements of OTU Picking Method

Supplementary Material 2 - Diversity Measurements for OTU picking methods



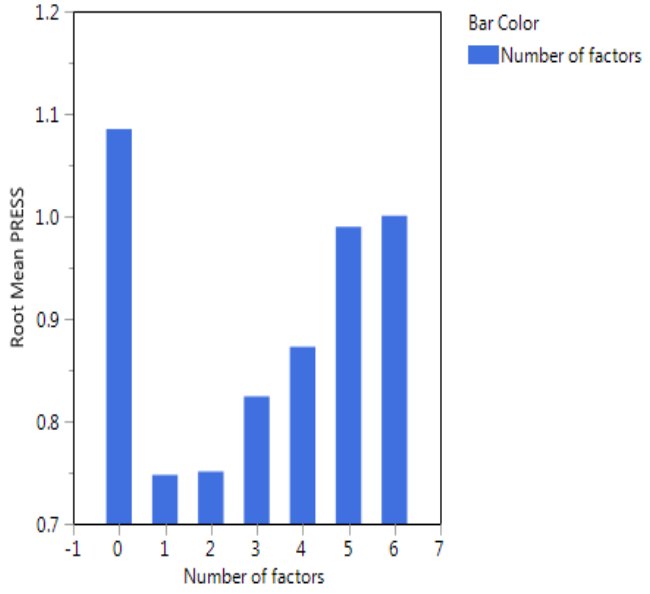
Supplementary 3: PLSR Validation and Diagnostic Plots

Fresh Water Model: 2nd run with VIP variables

KFold Cross Validation with K= 7 and Method=NIPALS

Number of factors	Root Mean PRESS	van der Voet T^2	Prob > van der Voet T^2	Q^2	Cumulative Q^2	R^2X	Cumulative R^2X	R^2Y	Cumulative R^2Y
0	1.085168	5.464737	0.0020*	-0.202395	-0.202395	0.000000	0.000000	0.000000	0.000000
1	0.747657	0.000000	1.0000	0.429234	0.429234	0.471727	0.471727	0.613760	0.613760
2	0.751165	0.002584	0.9700	0.423865	0.671162	0.164465	0.636192	0.083966	0.697725
3	0.824462	0.628532	0.4580	0.305943	0.771768	0.081458	0.717650	0.029054	0.726779
4	0.872536	0.971622	0.3670	0.222644	0.822582	0.119690	0.837340	0.001913	0.728693
5	0.989820	1.482887	0.2880	-0.000381	0.822515	0.096370	0.933710	0.000959	0.729651
6	1.000585	1.492837	0.2870	-0.022260	0.818564	0.066290	1.000000	0.000096	0.729747

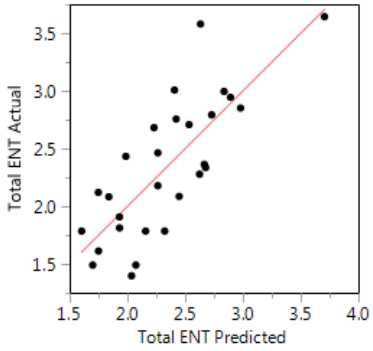
Root Mean PRESS Plot



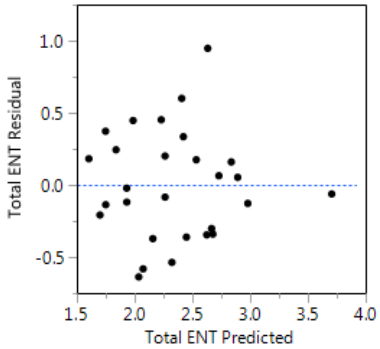
Note: The minimum root mean PRESS is 0.74766 and the minimizing number of factors is 1.

Diagnostics Plots

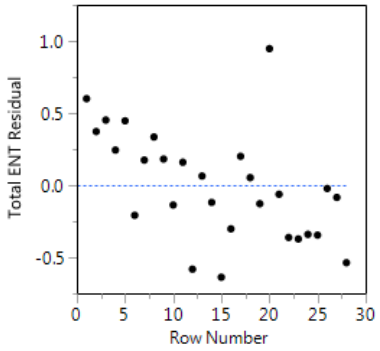
Actual by Predicted Plot



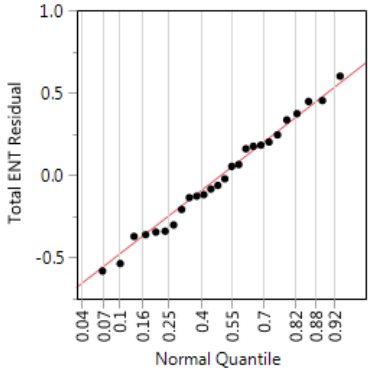
Residual by Predicted Plot



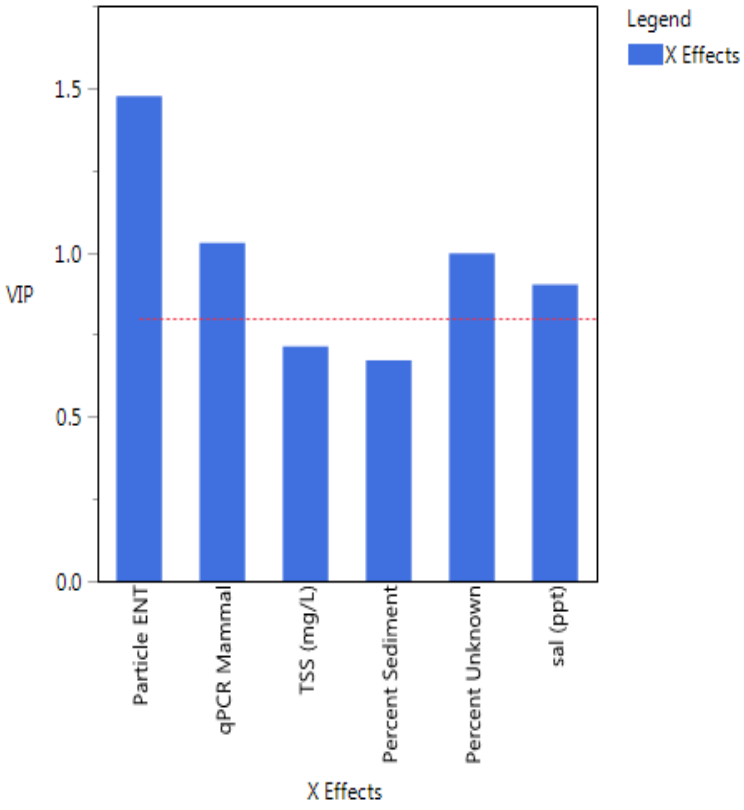
Residual by Row Plot



Residual Normal Quantile Plot



Variable Importance Plot



Variable Importance Table

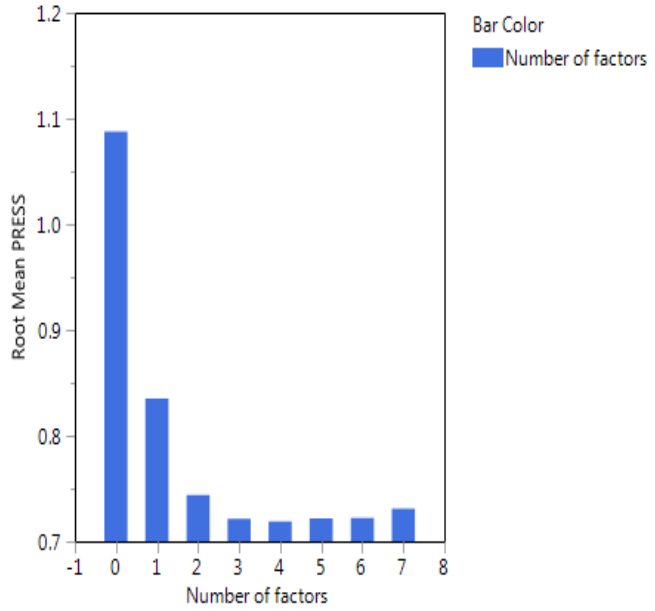
X	VIP
Particle ENT	1.4752
qPCR Mammal	1.0281
TSS (mg/L)	0.7134
Percent Sediment	0.6717
Percent Unknown	0.9961
sal (ppt)	0.9024

Estuary, Estuary Beach, and Marine Beach Model: 2nd run with VIP variables

KFold Cross Validation with K= 7 and Method=NIPALS

Number of factors	Root Mean PRESS	van der Voet T^2	Prob > van der Voet T^2	Q^2	Cumulative Q^2	R^2X	Cumulative R^2X	R^2Y	Cumulative R^2Y
0	1.087776	19.705955	<.0001*	-0.121733	-0.121733	0.000000	0.000000	0.000000	0.000000
1	0.835245	4.180911	0.0300*	0.336526	0.336526	0.316345	0.316345	0.481269	0.481269
2	0.743981	1.005251	0.3320	0.472614	0.650093	0.246863	0.563208	0.143132	0.624400
3	0.721602	0.042318	0.8430	0.505261	0.826887	0.252361	0.815569	0.030045	0.654445
4	0.719038	0.000000	1.0000	0.508137	0.914852	0.116803	0.932373	0.004762	0.659207
5	0.721937	0.370232	0.6400	0.503650	0.957737	0.034432	0.966804	0.000952	0.660159
6	0.722254	0.423888	0.6140	0.503192	0.979003	0.033196	1.000000	0.000012	0.660171
7	0.730902	3.885143	0.0380*	0.491130	0.989315	0.000000	1.000000	0.001885	0.662056

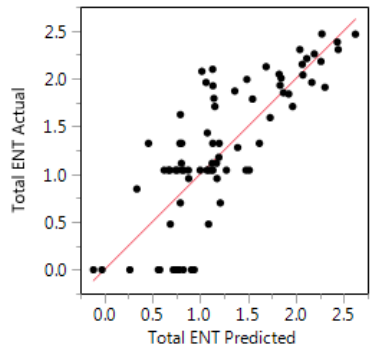
Root Mean PRESS Plot



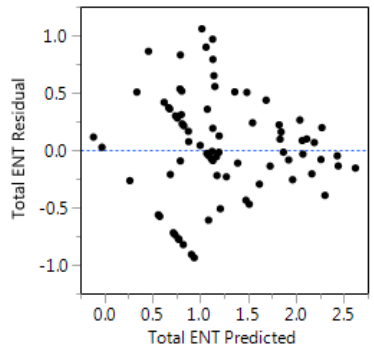
Note: The minimum root mean PRESS is 0.71904 and the minimizing number of factors is 4.

Diagnostics Plots

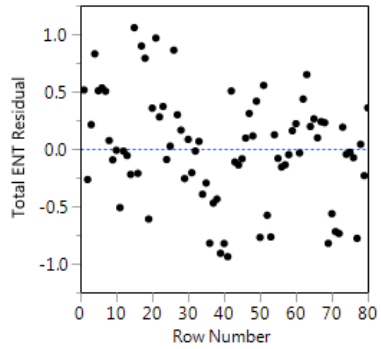
Actual by Predicted Plot



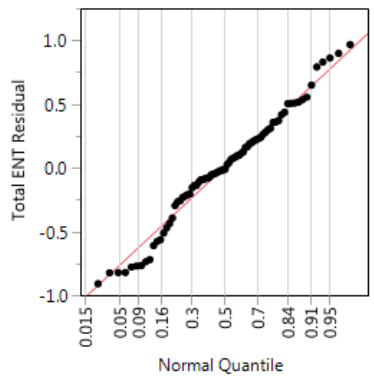
Residual by Predicted Plot



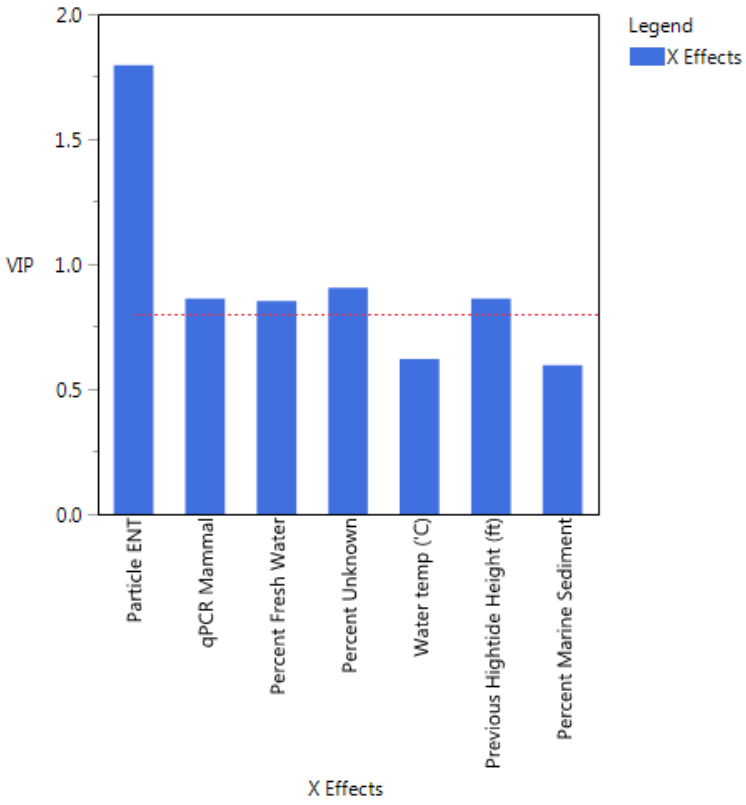
Residual by Row Plot



Residual Normal Quantile Plot



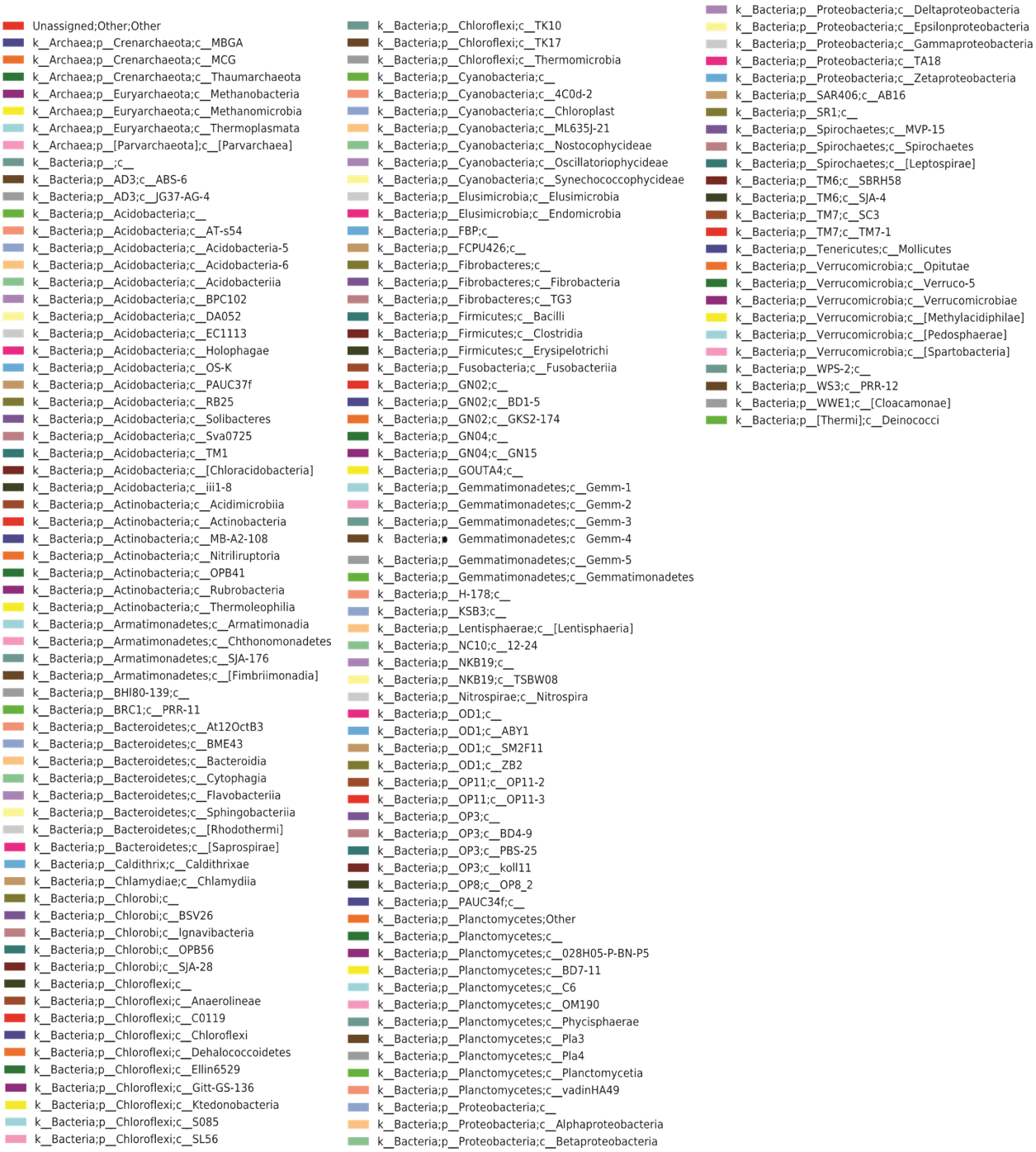
Variable Importance Plot



Variable Importance Table

X	VIP
Particle ENT	1.7954
qPCR Mammal	0.8621
Percent Fresh Water	0.8531
Percent Unknown	0.9066
Water temp (°C)	0.6203
Previous Hightide Height (ft)	0.8622
Percent Marine Sediment	0.5962

Supplementary Material 4: Taxa Key for all Class-level Assignments.



Chapter 3: Abundance of Fecal-Borne Potential Pathogenic Bacteria and Their Relationship to Fecal Sources in Three Maine Coastal Watersheds

Abstract:

Coastal waters worldwide have been influenced by human activities for centuries, with one consequence being increased fecal pollution of coastal waters via point and non-point source contamination. Pollution of coastal waters from fecal sources is a significant public health concern due to the wide arrays of fecal-borne pathogens that can make humans sick. Microbial source tracking (MST) methods allow for detection and quantification of the fecal sources present in coastal waters, however their significance as potential sources of bacterial pathogens is not well characterized. In this study, we analyzed 238 surface water samples from study areas in three separate coastal watersheds in southern Maine, USA. We used enterococci MPN analyses, MST qPCR assays targeting mammals, humans, and birds to quantify predominant fecal sources and 16S rRNA gene sequencing to identify total and fecal-specific potential pathogenic bacteria. The results indicated that water samples from urban storm drains contained the most total and fecal specific potential pathogenic bacteria. There were no significant relationships between total potential pathogens and any specific fecal source, however the total potential pathogens included *Vibrio* species and other marine-related and other non-fecal associated species. There were significant positive correlations between concentrations of both enterococci ($r^2 = 0.25$, $p < 0.05$) and bird fecal markers ($r^2 = 0.27$, $p < 0.05$) with the percent of fecal-specific potential pathogens detected. Overall, these results suggest that storm water and birds may be significant sources of fecal-borne bacterial pathogens. Types of water and fecal sources may vary in public health significance and both are important factors for evaluating the presence and detection of potential pathogens in coastal recreational water quality.

1. Introduction:

Coastal recreational water quality has been a long-term and significant issue related to both public health in keeping recreational waters clean, and the potentially significant economic implications for local communities if poor water quality is associated with their area. To regulate these waters the EPA has established guidelines that rely on fecal indicator bacteria, specifically enterococci, to represent fecal-borne contamination and potential public health risks in coastal recreational waters. Enterococci, however, are not exclusive to animal fecal tracts and they are capable of existing and persisting in aquatic ecosystems (47). This can confound the meaning of measured enterococci concentrations in recreational waters, potentially leading to elevated risks being associated with waters that might pose little to no risk to the public (53).

Microbial source tracking (MST) methods have helped delineate fecal sources that are present in recreational waters, however limited work has been done to evaluate the range of source-specific MST markers and how their concentrations within the environment correspond to public health risk. Work conducted by Boehm et al. (107) determined 4,200 copies of the human fecal source assay (HF183) was the cutoff above which there are >30 predicted GI illnesses/1000 swimmers, however similar studies have not been conducted for other assays. Other studies have related MST source marker concentrations to selected bacterial and viral pathogen levels with mixed results. Some studies have found significant relationships between *E. coli* O157:H7, *Salmonella* spp., *Campylobacter jejuni* and human fecal source (HF183; 141, 142) while others have found

no significant relationships (113, 143). These studies also point out the limitation of using a few select pathogens to investigate the public health significance of fecal pollution indicated by MST source marker detection.

Next generation sequencing (NGS) has recently become more robust and affordable, allowing for deeper sequencing of the 16S rRNA gene. Recent MST studies have used NGS to fingerprint bacterial communities from different source material (89) and to identify and relatively quantify potential human pathogens (29, 151, 152). Other MST studies that have used NGS have confirmed that different land-uses and wastewater treatment facilities and associated infrastructure can influence potential pathogen abundance in recreational waters (151–153). The use of NGS technologies to investigate potential pathogen abundance is more robust compared to other approaches in the sense that, theoretically, the full array of potential bacterial pathogens can be identified rather than using select pathogens as indicators (86). This allows for site-specific potential pathogen profiles to be identified and related to indicator levels and other measures according to public health significance.

In this study, we sought to identify relationships between the abundance of potential pathogen taxa using NGS sequencing and MST fecal marker copy numbers for mammals (Bac32), humans (HF183), and birds (GFD), along with enterococci concentrations. We analyzed a robust database of 238 surface water samples from field studies conducted at three popular recreational marine beach watersheds in Southern Maine, USA across two years. Data for enterococci concentrations and MST fecal marker concentrations were clustered and resulted in four unique

groups with different patterns for concentrations of markers and enterococci that had different abundance and diversity of potential pathogens identified. The results showed potential pathogen presence is greater when enterococci and MST fecal marker concentrations are elevated, although this is also influenced by location and sample type, and that storm water and bird fecal sources may be important sources of bacterial pathogens in these areas.

2. Methods:

2.1 Study Sites. Three different areas in Southern Maine were targeted in this study, all of which feature popular coastal beaches. The Wells, ME area includes an estuary with several beaches, two major freshwater inputs, and a large coastal beach. These beaches have historically had elevated levels of enterococci and are a concern in Maine (MHB 2015).

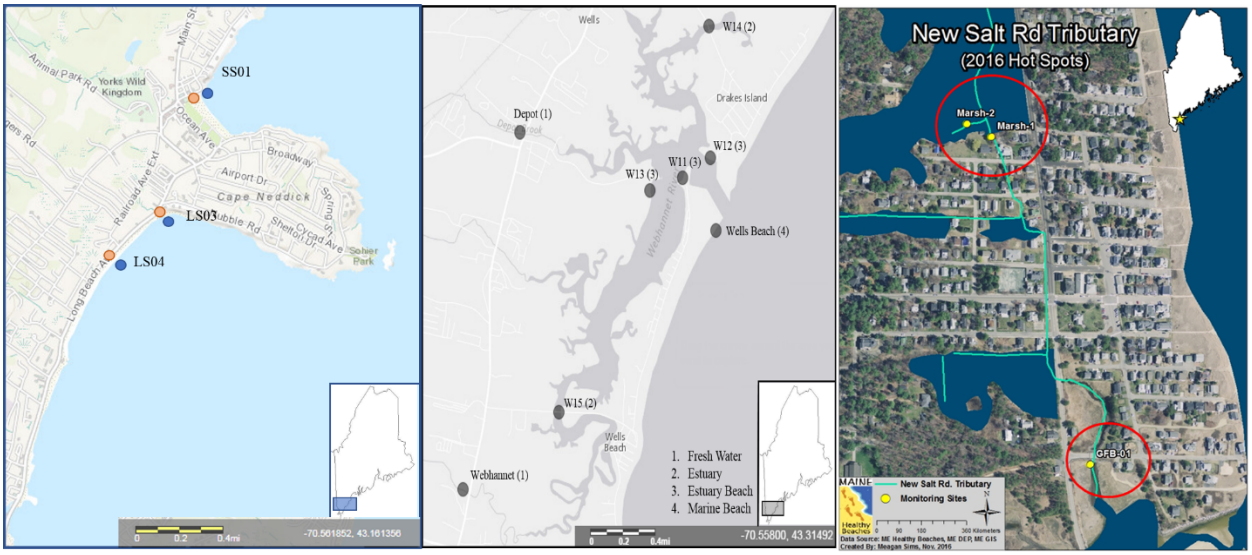


Figure 1: Study site locations for the York (2015), Wells (2016), and Old Orchard Beach (2016) studies. York (left), sample locations are represented by circles, with orange corresponding to storm drains and blue to beach water. Wells (middle), sample locations are designated with a black circle with numbers corresponding to water type. Old Orchard Beach (right) sample areas are circles in red, with sampling site denoted by a yellow circle, and the blue color within the urban area indicates a wetland.

In 2016 we conducted a MST study on the role of fecal sources, environmental conditions, and environmental source contributions in influencing enterococci concentrations *(In prep)*. York, ME has four major coastal beaches, two of which have untreated storm water discharges onto the beaches that drain directly into the ocean (Figure 1). York is currently addressing this condition according to Municipal Separate Storm and Sewer (MS4) regulations in collaboration with the USEPA. Old Orchard Beach, ME (OOB) is arguably the most popular beach in Maine, and for years the area has had problems with elevated enterococci levels in a marsh area and a tidal river that discharge directly into the beach area. Since 2016 we have partnered with the Maine Healthy Beaches Program to help identify both fecal sources and potential hotspots of fecal contamination in those areas.

2.2 Sample Collection. Surface water samples were collected during the summer tourist season, late May – early October in 2015 for York and in 2016 for the OOB study, and from June to September 2016 for the Wells study. A total of 238 water samples were collected across these three studies, with 108 for Wells, 107 for York, and 23 for Old Orchard Beach. All samples were collected during low tide and before noon. Water samples were collected in either autoclaved 1L Nalgene Wide-Mouth Lab Quality PPCO bottles (Thermo Fisher Scientific, Waltham, MA, USA) or 500 ml Whirl Pak™ sample collection bags (Nasco, Atkinson, WI, USA).

2.3 Enterococcus Quantification and Sample Processing. Enterococci quantification was performed using the Enterolert MPN assay (IDEXX, Westbrook, ME, USA), which is an

approved ASTM method (#D6503-99) for recreational water quality. For DNA work, 500 ml volumes for each sample were filtered through a 0.45 µm Nitrocelluose membrane (Millipore™, Darmstadt, Germany), placed into a cryovial tube and stored at -80˚C until DNA extraction.

2.4 DNA Extractions. All DNA extractions were performed using the PowerSoils® DNA extraction Kit (MO BIO Laboratories, Carlsbad, CA, USA), with modifications to the manufacture's protocol to optimize the extraction from the filter. Prior to DNA extraction, frozen filters were crushed into small pieces with an ethanol sterilized razor blade, a practice commonly used to maximize DNA recovery (123, 124). To minimize additional DNA loss during the extraction process solutions C2 and C3 (from manufacture's protocol) were halved in volume and combined into a single step.

2.5 MST qPCR. Quantitative PCR assays for MST source markers were run to determine fecal source strength for Mammals (AllBac, 104), Humans (HF183, 29), and Birds (GFD, 40). All qPCR assays were run on a Mx3000P cycler (Agilent Technologies, Santa Clara, CA, USA), TaqMan assays used the PerfecCTa® FastMix® II (QuantaBio, Beverly, MA, USA) master mix and the SYBR green assay used the FastSYBR™ Green Master Mix (Applied Biosystems, Foster City, CA, USA). Positive control plasmids were made for each assay from an assay-specific fresh fecal sample using the TOPO™ TA™ Cloning Kits (Invitrogen, Carlsbad, CA, USA), with a blue/white screen of constructs on kanamycin (50 µg/mL) selective TSA plates. White colonies determined to be positives with their respective PCR assay were then grown in TSB and extracted with the PureLink® Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA). A

standard curve ranging from 10^6-10^1 copies was also run for each experimental run. The Ct values, amplification efficiency, slope, and R^2 values for each standard curve were compared to previously run standard curves, to ensure satisfactory performance before being used to calculate copies numbers for that run. Each environmental sample was diluted 1:10 and run in triplicate, the reaction volume was 25 µl and contained a final concentration of 0.2 mg/ml BSA. Amplification/cycling conditions were preformed per published protocols for AllBac (129), HF183 (130), and GFD (85) source-specific markers. TaqMan assays were run with an internal amplification control (131) with a down-shift of 1 cycle considered inhibition. Samples spiked with a plasmid containing 10^4 copies of GFD amplicon were used as inhibition controls for the SYBR assay, with a recovery of less than 10^4 copies (100%) considered inhibition.

2.6 16S Library Preparations. The V4 region of the 16S rRNA gene, using the 515F-806R primer-barcode pairs, was used for amplicon sequencing (133). The Earth Microbiome Project protocol was used for amplification and pooling of samples, with minor modifications. The Qubit® dsDNA HS assay was used to quantify sample concentrations, and 500 ng of DNA was pooled per sample. The pool was then run on a 1.2 % low-melt agarose gel to separate primer-dimers from acceptable product, and bands between 300-350 bps were cut and extracted as previously described. The final product was then run on the Agilent Technologies 2200 TapeStation system (Santa Clara, CA, USA) to determine final size, quality, and purity of sample. Each library was sent to the Hubbard Center for Genome Studies at the University of New Hampshire to be sequenced (2 x 250 bp) on the Illumina HiSeq 2000 (San Diego, CA, USA).

2.8 OTU Filtering. QIIME 1.9.1 was used to perform all major quality filtering, and OTU picking (Caporaso et al., 2010). Forward and reversed reads were quality trimmed (μ P25) and removed of Illumina adapters via Trimmomatic (135). Any reads that were less than 200 bps were discarded, and reads were merged with the QIIME joined_paired_ends.py, using a minimum overlap of 10 bps and max percent difference of 10%. Paired-end data were analyzed using the QIIME open-reference OTU picking strategy with UCLUST for *de novo* picking and the Greengenes 13_8 database (136) for taxonomic assignment. Data for all samples are available on the Short Read Archive (http://www.ncbi.nlm.nih.gov/sra/) project reference (_____).

2.9 Potential Pathogenic Taxa Screen. A database comprised of 538 known human pathogenic bacterial species was downloaded from NCBI. These strains were chosen as they have known human virulence factors within their genomes, as first reported by Cai and Zhang (29). A custom Python script was written to download the genome fna and gfp files from NCBI, then full-length 16S rRNA sequences were extracted from gfp files using a regular expression to parse for the 16S rRNA gene identifier. For information on bacterial species used for the pathogenic database refer to Supplementary Material 1A. From the 538 total potential pathogens, a fecal-specific potential pathogenic data was constructed containing 83 different bacteria, which was based off of previous studies (Supplementary Material 1B; 4, 10, 27, 28). Sequencing data was then binned based on the cluster analysis of enterococci and qPCR data, resulting in four groups of data. Each group contained raw sequences (seqs.fna) and was subjected to a BLASTn search against both the constructed total potential pathogenic database and the fecal-specific database. A positive

identification was determined if the match was greater than 200 bp, 99% identical, with an e-value cutoff of 1 x 10^{-20} (156).

2.10 Statistical Analysis and Visualizations. All routine statistical analyses were performed in either R v3.4.0 or Python 3.6.1. Multivariate analyses were performed with PC-ORD v6. Graphing was performed in IPython notebook with matplotlib, seaborn, pandas, and numpy packages. All pairwise comparisons were done using the Kruskal-Wallis nonparametric method, with Dunn's nonparametric multiple comparisons run *post hoc* using a Bonferroni correction.

3 Results

The goal of this study was to determine if the concentrations of different sources of fecal contamination correspond to potential pathogen abundance. To investigate this, we chose to analyze the data via two different strategies. First, we clustered the data based on enterococci and MST marker concentrations, independent of location to determine higher-level groupings of these data and to allow for identification of relationships with potential pathogen abundance. Our second analysis investigated spatial trends to determine if those relationships were specific to location or water-type.

3.1 Site Specific Enterococci MPNs and MST Fecal Source Concentrations. We measured enterococci and MST marker concentrations for mammal, human, and bird fecal sources to characterize the relative fecal contamination across all sites and studies. There were patterns apparent for enterococci and marker copy number concentrations in water matrices (i.e. freshwater, storm drain, estuary) and study site (Figure 2). Enterococci showed significant variation within study, with the statistically significant highest average concentrations at the

OOB marsh sites (2.7 x 10^3 MPN/100 ml), followed by York storm drains (3.2 x 10^2 MPN/100 ml) and Wells freshwater sites (2.1 x 10^2 MPN/100 ml; $p < 0.05$). The mammal marker concentrations for the Wells study showed significant differences between higher concentrations in the estuary (1.1 x 10^7 copies/100 ml) and lower concentrations in freshwater and marine beach water (1.9 x 10^6 copies/100 ml), while the York study also showed significant differences between the higher mammal marker concentrations in storm drains (6.7 x 10^5 copies/100mLs) compared to lower concentrations in beach water (1.9 x 10^4 copies/100mLs; $p < 0.05$).

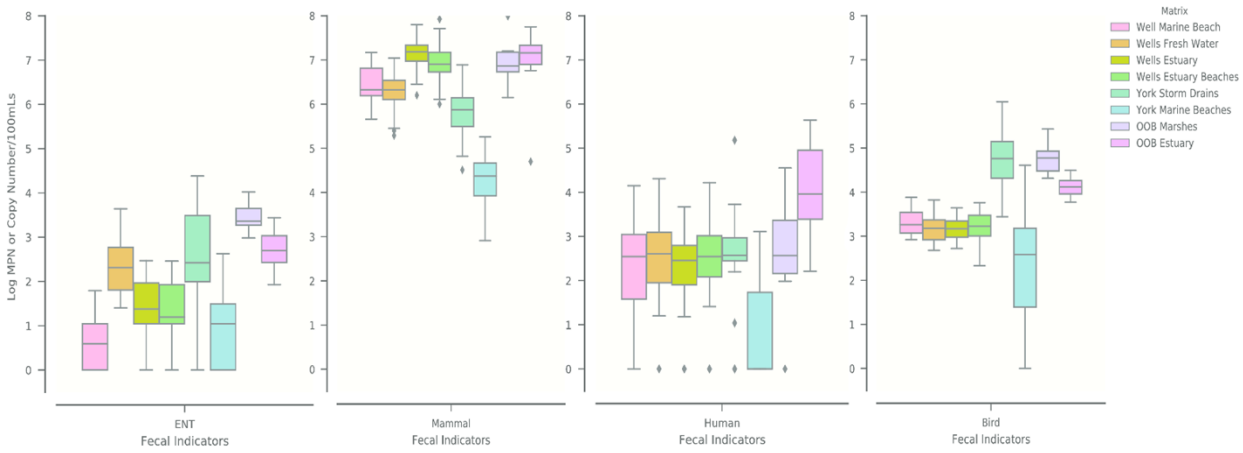


Figure 2: Enterococci (MPN/100 ml) and Fecal Source (copy number/100 ml) Concentrations in Water from Wells, York, and Old Orchard Beach study areas. Minimum and maximum data points are depicted at the edge of the box plot whiskers, with outliers denoted with dots. The edges of the boxplots correspond to the first and third quartiles, and the middle represents the median of the data.

The human marker concentrations showed a great degree of variation within each specific site, but average concentrations were not statistically different for the Wells study sites. However, copy number concentrations in storm drains and beach water (York study) and marsh and tidal river-GFB01 (OOB study) were statistically different from each other (Figure 2, $p < 0.05$). The bird fecal marker concentrations were very consistent at all Wells sites, (1.6 x 10^3 copies/100 ml ± 1.9 x 10^0). The York study showed significantly elevated bird marker copy numbers in the

storm drains compared to beach water ($p < 0.05$). Concentrations of bird fecal markers differed significantly for the OOB study, with levels significant higher in the marsh (5.8×10^4 copies/100 ml) compared to the tidal river-GFB01 (1.3×10^4 copies/100 ml; $p < 0.05$). Overall, results showed different characteristics with regards to inter-study site-to-site variation. The Wells study showed significant difference between sites for enterococci and mammal marker concentrations, while the York and OOB studies showed differences between sites for enterococci and all fecal markers.

3.2 Between Study Differences in Enterococci and MST Fecal Source Concentrations. Although there were many significant differences within each study, there were also differences between studies (Figure 2). When compared to all other studies, OOB contained the highest concentrations of enterococci (1.1×10^3 MPN/100mLs), and MST markers (mammal = 8.3×10^6, human = 1.5×10^3, and bird = 3.0×10^4 copies/100mLs), all of which were significantly higher ($p < 0.05$) than at other study areas, except the mammal fecal marker compared to the Wells study. Storm drains in York and freshwater at Wells sites showed similar numbers for enterococci and MST markers, with the exception that the bird fecal marker was significantly higher in storms drains (Figure 2; $p < 0.05$). When beach water analysis results from Wells and York studies were compared, fecal source markers for mammal and humans were statistically higher for Wells (Figure 2; $p < 0.05$), but enterococci concentrations were similar between studies. Together with intra-study site-to-site variation, between studies showed a similar degree of significant differences indicating unique patterns of enterococci and fecal source markers between site locations and across studies.

3.3 Grouped Data Based on Enterococci and MST Concentrations. Our first analysis strategy sought to investigate if there was a general relationship between fecal marker concentrations and potential pathogen abundance, independent of location. To achieve this, we chose to cluster enterococci and MST marker concentrations using a Ward's cluster analysis with Euclidean distance. Results showed four distinct groups when 60% of the data was used to resolve independent groups. This was chosen because the PCORD software deemed it to be the ideal solution based on the Calinski & Harabasz index (Figure 3).

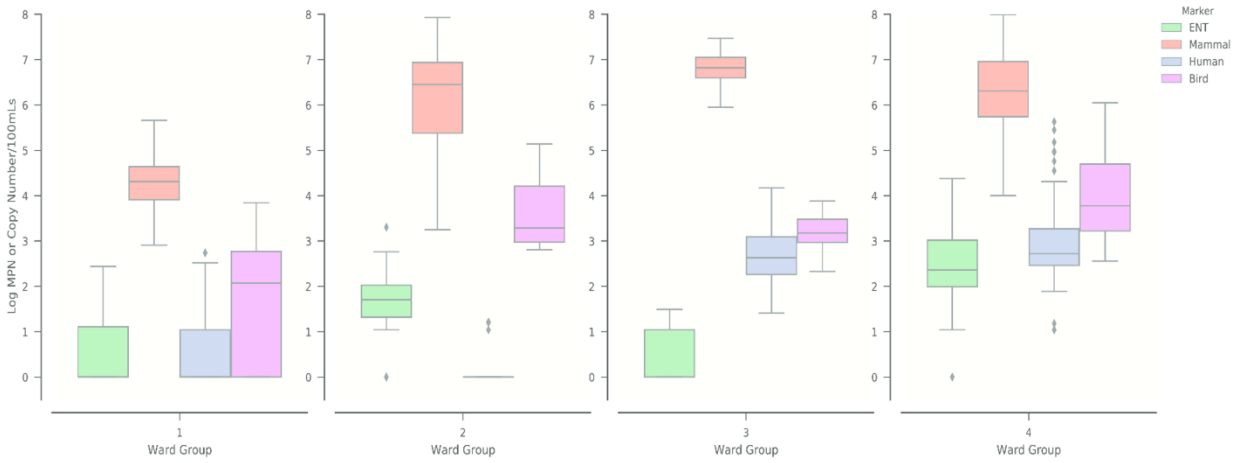


Figure 3: Enterococci and MST qPCR Concentrations for Ward Clustered Groups. Results shown are concentrations for enterococci and MST fecal markers for each clustered dataset. Colors correspond to enterococci (green), mammal (orange), human (blue), and bird (pink) concentrations.

Each clustered dataset showed a unique pattern that was statistical different ($p < 0.05$); where Group 1 contained low levels of enterococci and MST markers, Groups 2 & 3 were not different except for enterococci and human concentrations, and Group 4 contained elevated levels of enterococci and MST markers (Figure 3). The resulting groups of data enabled downstream analysis for potential pathogen abundance and relationships with MST marker concentrations.

3.4 16S Profiles of MST Clusters. A total of 3,174,326 reads were obtained from all datasets, which resulted in 28,712 unique OTUs. Most assigned taxa were Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, and Flavobacteriia for each clustered dataset (Figure 4).

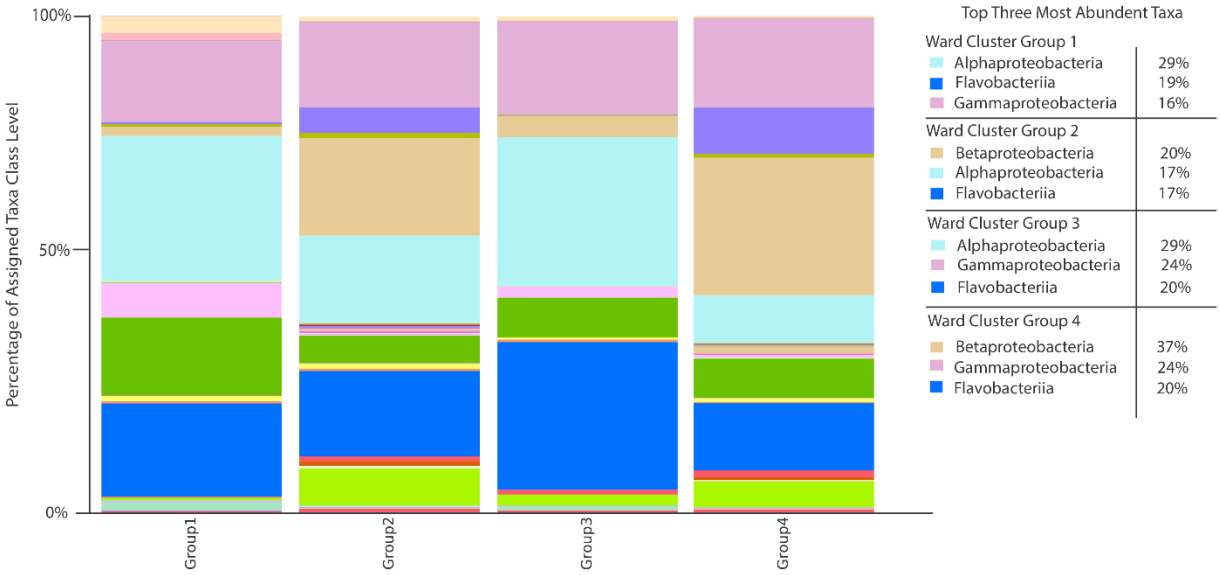


Figure 4: 16S Community Composition Relatively Similar Between Clustered Sample Groups. Shown is the 16S community composition for each clustered dataset. A key to the right indicates the most abundant classes for each dataset.

Taxa in each clustered sample group generally consisted of the same types of bacterial classes. The top 4 classes in each cluster were Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, and Flavobacteriia. However, the relative abundance of each class differed between group datasets. Cluster Group 4, which had high levels of enterococci and MST

fecal markers, had a significantly higher amount of Betaproteobacteria and significantly lower amount of Alphaproteobacteria when compared to cluster groups. Cluster Groups 1 & 3 contained very similar compositions, with Alphaproteobacteria, Flavobacteriia*, and* Gammaproteobacteria all at similar proportions.

3.5 Abundance of Potential Pathogens within Clustered Datasets and Between Studies. To determine the relationship between potential pathogen presence and concentrations of enterococci and fecal markers, we constructed a 16S rRNA blast against a constructed potential pathogen database (Methods 2.9). Results showed cluster Group 4 contained the highest levels of potential pathogens identified (5.4%), 34% of which belonged to *Vibrio fluvialis* (Table 1). Cluster Group 1, which had the lowest levels of enterococci and MST markers contained the second highest levels of potential pathogens (3.8%), however, 80% were identified as *Vibrio parahaemolyticus*. Out of all four groups, cluster Group 3 showed the lowest number of potential pathogens (0.09%), and consequently only 16 potential pathogens were identified. On the other hand, Group 4 showed significantly higher levels of potential pathogens identified. This was also reflective in the Shannon's Diversity Index, as Group 4 was significantly higher compared to other groups, indicating greater diversity of identified potential pathogens (Table 1, Figure 5).

Cluster MST Group	Percent of Potential Pathogens	Average Percent ID	Top Potential Pathogen	Total Potential Pathogens ID'd	Shannon's Diversity Index (H)
1	3.83%	99.66%	*Vibrio parahaemolyticus (79.67%)*	73	0.931
2	1%	99.67%	*Escherichia coli (43.53%)*	77	1.529
3	0.09%	99.63%	*Vibrio alginolyticus (50.66%)*	16	1.649
4	5.39%	99.64%	*Vibrio fluvialis (33.91%)*	126	2.067

Table 1: The Relative Abundance and Diversity of Potential Pathogens Identified in Clustered Sample Groups. Each row corresponds to a clustered dataset with columns showing key statistics from blast results.

Samples that contained >5% of potential pathogens (11% of the data) were analyzed further, and it was shown that all but one sample (Depot Brook 7-30-16) came from the York storm drains and beach water. On average, 67% ±39 of the total potential pathogens identified were a *Vibrio spp.* in these samples with greater than 5% of potential pathogens. Interestingly there was a large significant difference between the dominant *Vibrio* spp. identified in the storm drain vs. beach water. The storm drain contained a high percent of *V. fluvialis* where the beach sample waters contained a high percentage of *V. parahaemolyticus* (Supplementary Material 4; $p < 0.05$). Overall, the group of samples that contained higher levels of enterococci and MST fecal markers (cluster Group 4) also contained higher levels of potential pathogenic bacteria. Consequently, all but one sample that contained >5% of potential pathogens identified were from the York study, indicating a study specific pattern.

3.6 Abundance of Potential Pathogens and Fecal Specific between Studies. Given this study was conducted in Southern Maine, *Vibrio* spp. and others species within the 538 potential pathogen database are likely not fecal in origin. To account for this, the study data were re-screened using 83 fecal specific potential pathogens (Supplementary Table 1b) and then analyzed for spatial patterns., York had the highest percent of fecal specific potential pathogens, which was statistically higher than the Wells and OOB studies (Figure 6a; $p < 0.05$). Specifically, the storm drains at York showed the highest percent of fecal specific potential pathogens when

compared to all other types of sample locations (Figure 6b; $p < 0.05$).

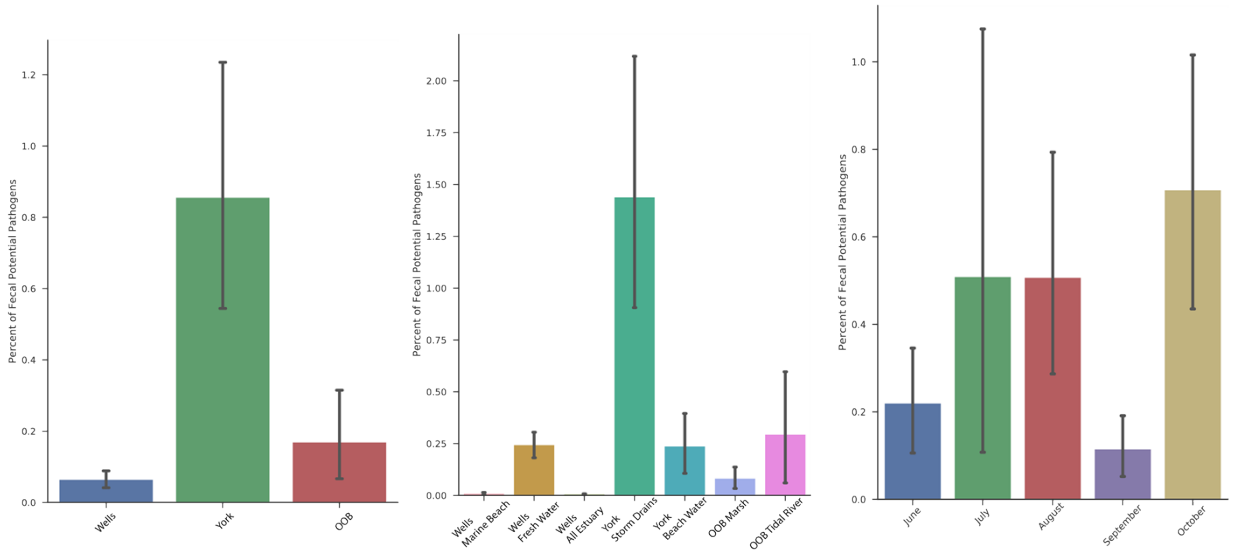


Figure 6: The Relative Abundance of Fecal Specific Potential Pathogens Between Study Areas (left), Sampling Sites (middle), and Months (right).

There were also significant differences between different months of the year and the percent of fecal specific potential pathogens identified; both August and October showed significantly higher levels of fecal potential pathogens compared to other months (Figure 6c; $p < 0.05$). Elevated levels of fecal potential pathogens were also observed for the month of July, but the data contained significant variation, making the increase not statistically significant.

3.7 Relationship Between Potential Pathogens and Enterococci and MST Concentrations. A key question for this study is whether concentrations of enterococci and fecal source-specific markers are related to the abundance of fecal-specific potential pathogen taxa. Relationships between fecal specific potential pathogens and enterococci and MST fecal markers were

investigated using linear regressions. For all the data, enterococci ($r^2 = 0.25$; $p < 0.05$) and bird fecal marker ($r^2 = 0.27$; $p < 0.05$) concentrations had significantly positive relationships to fecal-specific potential pathogens while there were no significant relationships with human and mammal fecal marker concentrations (Figure 7). These relationships were then analyzed by month to see if associations strengthen under the same months that saw significant levels of fecal specific potential pathogen increase. Results showed August and October had significantly positive relationships for enterococci and all fecal marker concentrations except mammal to percent of fecal specific potential pathogens (Supplementary Material 3).

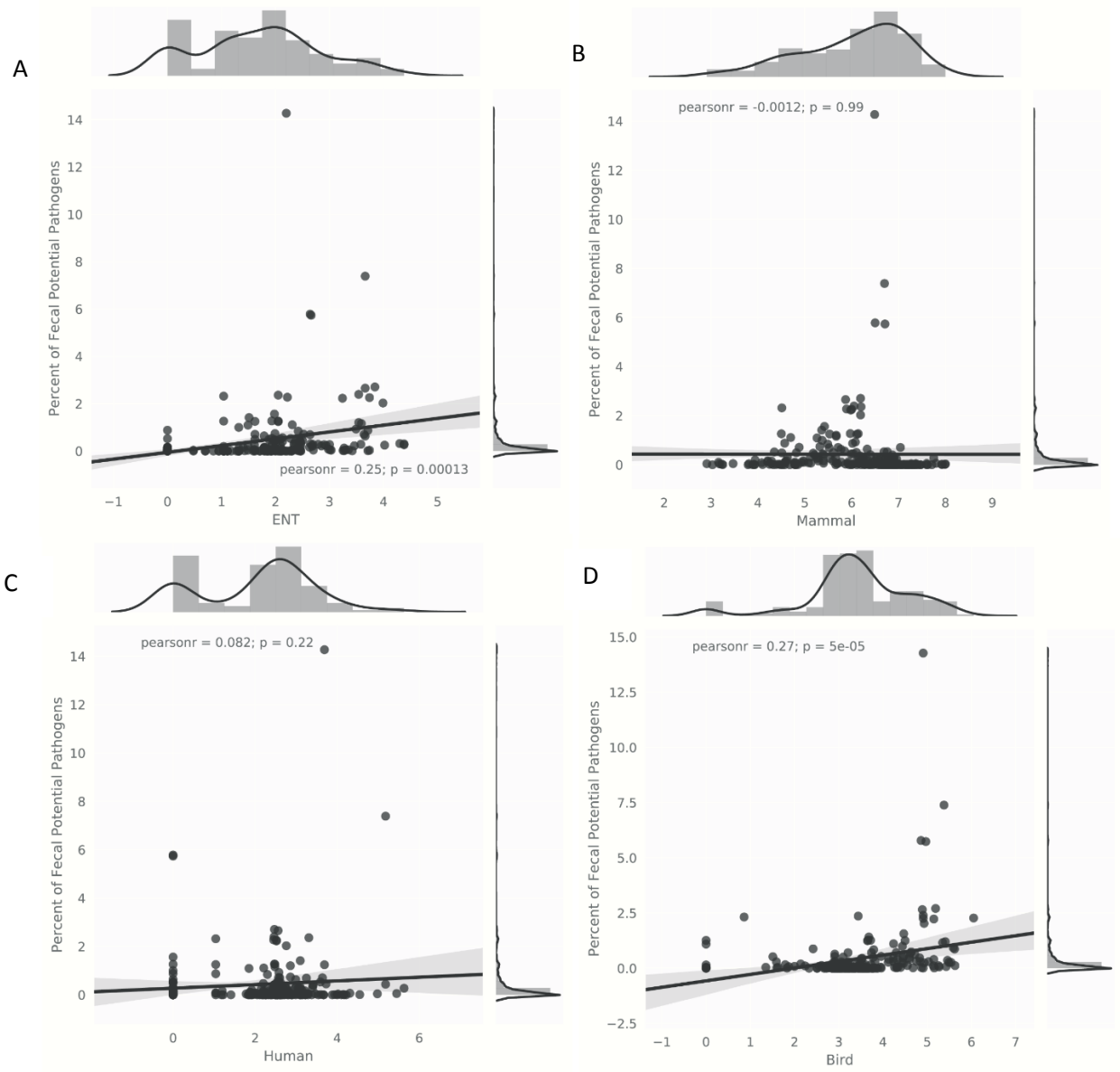


Figure 7: Linear Regressions of Fecal Specific Potential Pathogens with MST Fecal Marker and Enterococci Concentrations. (A), mammal fecal (B), human fecal (C), and bird fecal (D) concentrations. Histograms on the X and Y axes correspond to the distribution of the data for that specific axis.

For the human marker, these were the only two months where the relationship to fecal specific potential pathogens was significant (Supplementary Material 3). Out of all fecal markers, the bird concentrations showed significant relationships to fecal specific potential pathogens for all months, except September (June: $r^2 = 0.74$, July: $r^2 = 0.32$, August: $r^2 = 0.28$, and October: $r^2 = 0.44$; $p < 0.05$) with June significantly higher in its positive correlation ($p < 0.05$). Overall, results indicate that enterococci and bird fecal marker copy concentrations have the strongest relationship to fecal potential pathogens, with human marker showing a positive relationship in September and October.

4 Discussion

A key assumption for the use of fecal indicators is that their concentrations relate to the relative concentrations and presence of fecal-borne pathogens. In this study, we explored the use of DNA sequencing to identify taxa associated with potential pathogens and relate the percentages of these taxa within the microbial communities of samples to both enterococci and MST source specific marker copy number concentrations. Other studies have used similar approaches in trying to identify potential pathogenic taxa in wastewater treatment facilities (29, 151), and from a mixed urban watershed (152). Although to date, few studies have applied NGS technology for potential pathogen identification, so comparisons among studies are limited. This may be due to the fact that early applications of NGS for MST studies relied on pyrosequencing, which at the time lacked the ability to sequence at sufficient depth to be used for analyses such as the ones presented here (157). However, results from Ibekwe et al. (152) showed that the bacterial community in a wastewater treatment facility's discharge could contain up to 4.3% potential pathogen taxa, and that urban runoff could contain up to 7.9% of potential pathogen taxa. These

percentages are comparable to more recent work conducted in wastewater treatment facilities in China that also measured potential pathogen taxa abundances in facility effluent samples by Cai et al. (52). However, this study compares best to a recent work published by Li et al. (151) where a database of 538 different potential pathogens (same database as this study) was constructed and applied to wastewater treatment samples. Their findings showed that wastewater influent community taxa could contain up to 23.2% potential pathogen taxa and effluent could contain up to 11.8% potential pathogen taxa (151). Results presented here showed clustered Group 4 (elevated levels of ENT and MST markers) sample communities included 5.4% as potential pathogen taxa, which is likely a relatively elevated concentration to be observed in recreational waters. However, percentages of potential pathogen taxa were largely skewed by the identification of *Vibrio* spp., which is why we chose to further analysis the data by only looking at fecal specific potential pathogens. This finding is highly probable given that *Vibrio* spp. are natural inhabitants of coastal waters (158), yet are not generally fecal-borne in the U.S. To our knowledge there is no similar study that has investigated fecal-specific potential pathogens using similar methodologies.

In the MST field, human fecal contamination represents the greatest public health risk when compared to other fecal source contamination (8, 9, 25, 143). This is largely due to the "no-species" barrier as well as the higher diversity of pathogens that humans can carry. One goal of this study was to investigate whether different patterns in fecal source concentrations obtained from the field might represent a difference in potential pathogens. Because our question revolved around this issue we chose to cluster data based on log-transformed MPN and MST copy number concentrations, independent of location sampled. Results showed that samples in clustered

Group 4 had significantly higher fractions of total and fecal specific potential pathogen taxa when compared to other groups. Within this group, the average ENT concentrations were 2.6 x 10^2 MPN/100 ml, which is above the EPA water quality limit (3). Also the average human marker concentration was at 5.8 x 10^2 copies/100 ml, which is close to the 4,200 copies/100 ml (HF183) cutoff determined by Boehm et al. (107), thus, signifying human contamination was at a level relatively close to the 30 predicted GI illnesses/1000 swimmers. These results indicate that elevated levels of enterococci and fecal markers can correspond to increased presence of potential pathogen taxa.

However, the more significant result from the analysis of total and fecal specific potential pathogen abundance was observed based on the type of study location. The York and OOB study are areas of a higher urban index compared to Wells (measured by number of houses in a geographic area), but the York study in contrast to the OOB study, investigated storm drains that mainly handle urban runoff (159). The storm drains indeed contained the highest percent of identified fecal specific potential pathogen taxa, which suggests that the origin of the water sample (e.g. storm water, beach water, freshwater tributary) is more important than concentrations of enterococci or fecal source markers. This is likely due to the fact the storm drains handle urban runoff (increased impervious surface cover) and also may have illicit sewage connections, making the transport of human-specific and other sources of fecal contamination both more likely and much easier (160). Recent testing by the US EPA showed variable presence of chemical indicators of human discharge in storm drains (Leslie Hines-Town of York, personal communication). This is most pronounced when compared to the OOB study, where some of the highest levels of enterococci and MST fecal markers were measured, but where there were

significantly lower fecal specific potential pathogen taxa. The sites in the OOB study were a marsh area and tidal river, both of which are very different from urban storm water. In both the marsh and tidal river there were elevated levels of enterococci and MST markers, in particular the tidal river contained the highest average of human marker copies, However given the nature of a marsh/tidal river ecosystem there is a high probability for potential pathogens to become sequestered within the sediments or flushed from the system. (161) So areas such as marshes or tidal rivers might be able to withstand higher fecal loads without higher levels of potential pathogens observed in the water when compared to urban stormwater runoff (47). Although fecal contamination is a source of pathogens to recreational waters, the persistence and survival of actual pathogens in the environment is controlled by variety of factors, including but not limited to environmental conditions (e.g. temperature, nutrient availability), predation from phage and bacterivores, competition, and other ecologically relevant factors that influence growth (52, 149, 162). Storm drains represent an environmental reservoir that may harbor more recent fecal contamination that has come more directly from sources, i.e., illicit sewage connections.

Enterococci and bird fecal marker concentrations were the only source-specific markers in this study where concentration showed a significant positive relationship to percent of fecal specific potential pathogen taxa. This finding for the bird fecal marker could be distinctive of coastal recreational areas, as these areas harbor large gull populations that are actively interacting with the beach and estuarine water, and thus the proximity of the fecal contamination is more direct compared to other fecal sources. Seasonal analysis also revealed that spikes in fecal specific potential pathogens also resulted in significant positive correlations with human fecal marker concentrations, indicated human-related sources as the reason for the seasonal trend. Other

studies investigating enterococci and selective pathogen abundance have found positive correlations between *C. perfringens* (11) and *Salmonella* spp. (114); however, a total of 86 fecal specific potential pathogens were used in this analysis, signifying a much broader depth in potential pathogens that could be identified. The detection of *E. coli* was the overwhelmingly dominant factor associated with the percentage of fecal specific potential pathogen taxa in this study. This is not surprising, as out of all fecal specific potential pathogen taxa *E. coli* is the most common species found in environment as well as the fecal tract (163). However, it might also highlight a potential limitation in amplicon sequencing to identify organisms that are at very low levels. In the final analysis, the approach still only can identify potentially pathogenic fecal-associated species and not actual pathogenic strains, so these results remain significant only as indications of the possible presence of pathogens and public health risks.

The World Health Organization (WHO) states bacterial pathogens account for 43% of the total pathogens that are of concern in drinking and recreational waters (164), meaning that over half of pathogens that are of concern are viruses, protozoans, helminths, fungi or algae. Studies investigating fecal contamination indicate that viruses (i.e. Norovirus) or pathogenic bacteria are the likely causative agents for most contaminated recreational waters (165, 166). EPA recreational water quality regulations, however, are designed around bacterial fecal indicators and the MST markers used in this study were also bacterial, making investigation of potential pathogenic bacteria logical for this study to enable relating findings to regulatory criteria. The results presented here, however, under-estimate the much wider array of actual pathogenic microorganisms and viruses that could be present. This study also used 16S rRNA to identify potential pathogens, yet it is well known that 16S rRNA does not directly reflect actual

pathogenicity for identified taxa. The best known example of this capacity is a study conducted by Welch et al. (167) where they showed that three *E. coli* strains (non-pathogenic, uropathogenic, and enterohaemorrhagic) that were 100% identical in 16S rRNA sequence but shared only 39% similar gene content. Programs exist that link phylogeny and functional content (168) and thus could be used to track pathogenicity traits, however, at best these programs can only infer core functions, and thus their use to predict pathogenicity or any horizontally transferred function would be misguided as presently designed. Given these considerations, the results presented in this study show interesting trends for the presence of potential pathogens and their relation to fecal contamination and location, and indicate potential public health risks associated with storm water discharges and both bird and human sources of fecal pollution.

The MST field is rapidly evolving, not only to develop new tools for tracking fecal pollution, but also to address how fecal pollution observed in field experiments might relate to actual public health risk. Results presented here showed that samples containing elevated levels of enterococci and MST fecal markers contained the highest percent of fecal specific potential pathogen taxa. However, most of those samples were from the York study, and, from storms drains. Thus, signifying study location and water matrix type are also important steps in framing the presence of potential pathogen taxa. Given that 16S rRNA is a slow-evolving gene, and could even be considered static when compared to the dynamic evolution of bacterial genomes, true pathogens cannot be identified using this approach. The dominance of *E. coli* amongst the fecal specific potential pathogen taxa could be the result of reduced fitness by other fecal specific potential pathogen taxa to survival in the environment, or they could just be at levels too low for detection. Further studies employing deep shotgun sequencing and proteomics could help identify true

pathogens and not simply potential pathogens, which could resolve fine scale relationships between different fecal sources and relative public health risks.

Supplementary Material 1A List of 583 Potential Pathogens

Abiotrophia defectiva	*Legionella oakridgensis*
Achromobacter piechaudii	*Legionella pneumophila*
Achromobacter xylosoxidans	*Legionella rubrilucens*
Acidaminococcus fermentans	*Legionella sainthelensi*
Acinetobacter baumannii	*Legionella tucsonensis*
Acinetobacter calcoaceticus	*Legionella wadsworthii*
Acinetobacter haemolyticus	*Leifsonia aquatic*
Acinetobacter johnsonii	*Leptospira borgpetersenii*
Acinetobacter junii	*Leptospira inadai*
Acinetobacter lwoffii	*Leptospira interrogans*
Acinetobacter radioresistens	*Leptospira kirschneri*
Actinobacillus actinomycetemcomitans	*Leptospira meyeri*
Actinobacillus hominis	*Leptospira noguchii*
Actinobacillus lignieresii	*Leptospira santarosai*
Actinobacillus equuli	*Leptospira weilii*
Actinobacillus pleuropneumoniae	*Leptotrichia buccalis*
Actinobacillus suis	*Listeria ivanovii*
Actinobacillus ureae	*Listeria monocytogenes*
Actinomyces radingae	*Listeria seeligeri*
Actinomyces georgiae	*Listeria welshimeri*
Actinomyces gerencseriae	*Mannheimia haemolytica*
Actinomyces israelii	*Megamonas hypermegale*
Actinomyces meyeri	*Megasphaera sp.*
Actinomyces naeslundii	*Micromonas micros*
Actinomyces neuii	*Mogibacterium timidum*
Actinomyces odontolyticus	*Moraxella (Branhamella) catarrhalis*
Actinomyces turicensis	*Moraxella (Branhamella) caviae*
Aerococcus viridans	*Moraxella (Branhamella) cuniculi*

Aeromonas caviae	*Moraxella (Branhamella) ovis*
Aeromonas hydrophila	*Moraxella (Moxarella) atlantae*
Aeromonas sobria	*Moraxella (Moxarella) bovis*
Aeromonas veronii	*Moraxella (Moxarella) lacunata*
Alcaligenes odorans	*Moraxella (Moxarella) liquefaciens*
Amycolatopsis orientalis	*Moraxella (Moxarella) nonliquefaciens*
Anaerococcus lactolyticus	*Moraxella (Moxarella) osloensis*
Anaerococcus prevotii	*Moraxella lincolnii*
Anaerococcus vaginalis	*Morganella morganii*
Anaplasma phagocytophila	*Mycobacterium abscessus*
Arcanobacterium bernardiae	*Mycobacterium africanum*
Arcanobacterium heamolyticum	*Mycobacterium asiaticum*
Arcanobacterium pyogenes	*Mycobacterium avium*
Arcobacter butzleri	*Mycobacterium bovis*
Arcobacter cryaerophilus	*Mycobacterium celatum*
Bacillus sphaericus	*Mycobacterium chelonae*
Bacillus anthracis	*Mycobacterium conspicuum*
Bacillus cereus	*Mycobacterium fortuitum*
Bacillus circulans	*Mycobacterium genavense*

Bacillus coagulans	*Mycobacterium gordonae*
Bacillus licheniformis	*Mycobacterium haemophilum*
Bacillus mycoides	*Mycobacterium kansasii*
Bacillus pumilus	*Mycobacterium leprae*
Bacillus subtilis	*Mycobacterium malmoense*
Bacillus thuringiensis	*Mycobacterium marinum*
Bacteroides distasonis	*Mycobacterium mucogenicum*
Bacteroides forsythus	*Mycobacterium peregrinum*
Bacteroides galacturonicus	*Mycobacterium porcinum*
Bacteroides merdae	*Mycobacterium scrofulaceum*
Bacteroides splanchnicus	*Mycobacterium senegalense*
Bacteroides ureolyticus	*Mycobacterium shimoidei*
Bacteroides caccae	*Mycobacterium simiae*
Bacteroides eggerthii	*Mycobacterium smegmatis*
Bacteroides fragilis	*Mycobacterium szulgai*

Bacteroides ovatus	*Mycobacterium tuberculosis*
Bacteroides pectinophilus	*Mycobacterium ulcerans*
Bacteroides stercoris	*Mycobacterium xenopi*
Bacteroides thetaiotaomicron	*Mycoplasma fermentans*
Bacteroides uniformis	*Mycoplasma genitalium*
Bacteroides vulgatus	*Mycoplasma hominis*
Bartonella bacilliformis	*Mycoplasma pneumoniae*
Bartonella elizabethae	*Mycoplasma salivarium*
Bartonella henselae	*Myroides odoratus*
Bartonella quintana	*Neisseria flava*
Bergeyella zoohelcum	*Neisseria cinerea*
Bifidobacterium dentium	*Neisseria elongata*
Bilophila wadsworthia	*Neisseria flavescens*
Bordetella avium	*Neisseria gonorrhoeae*
Bordetella bronchiseptica	*Neisseria lactamica*
Bordetella parapertussis	*Neisseria meningitidis*
Bordetella pertussis	*Neisseria mucosa*
Borrelia brasiliensis	*Neisseria perflava*
Borrelia caucasica	*Neisseria sicca*
Borrelia latyschewii	*Neisseria subflava*
Borrelia mazzottii	*Neisseria weaveri*
Borrelia venezuelensis	*Neorickettsia sennetsu*
Borrelia burgdorferi	*Nocardia caviae*
Borrelia crocidurae	*Nocardia asteroides*
Borrelia duttonii	*Nocardia brasiliensis*
Borrelia hermsii	*Nocardia farcinica*
Borrelia hispanica	*Nocardia nova*
Borrelia parkeri	*Nocardia otitidiscaviarum*
Borrelia persica	*Nocardia pseudobrasiliensis*
Borrelia recurrentis	*Nocardia transvalensis*
Borrelia turicatae	*Ochrobactrum anthropi*
Brevibacillus brevis	*Oligella ureolytica*

Brevundimonas diminuta	*Oligella urethralis*
Brevundimonas vesicularis	*Orientia tsutsugamushi*

Brucella melitensis	*Paenibacillus alvei*
Burkholderia cepacia	*Paenibacillus macerans*
Burkholderia mallei	*Pantoea agglomerans*
Burkholderia pseudomallei	*Pasteurella aerogenes*
Campylobacter coli	*Pasteurella caballi*
Campylobacter concisus	*Pasteurella canis*
Campylobacter curvus	*Pasteurella pneumotropica*
Campylobacter fetus	*Pasteurella stomatis*
Campylobacter gracilis	*Pasteurella dagmatis*
Campylobacter hyointestinalis	*Pasteurella multocida*
Campylobacter jejuni	*Peptinophilus asaccharolyticus*
Campylobacter lari	*Peptococcus niger*
Campylobacter rectus	*Peptostreptococcus anaerobius*
Campylobacter sputorum	*Photobacterium damselae*
Campylobacter upsaliensis	*Plesiomonas shigelloides*
Capnocytophaga canimorsus	*Porphyromonas asaccharolytica*
Capnocytophaga cynodegmi	*Porphyromonas catoniae*
Capnocytophaga gingivalis	*Porphyromonas circumdentaria*
Capnocytophaga ochracea	*Porphyromonas endodontalis*
Capnocytophaga sputigena	*Porphyromonas gingivalis*
Cardiobacterium hominis	*Porphyromonas levii*
Cedecea lapagei	*Porphyromonas macacae*
Cedecea davisae	*Prevotella heparinolytica*
Cedecea neteri	*Prevotella oulora*
Cellulomonas cellulans	*Prevotella tannerae*
Cellulomonas turbata	*Prevotella zoogleoformans*
Centipeda periodontii	*Prevotella bivia*
Chlamydia trachomatis	*Prevotella buccae*
Chlamydophila psittaci	*Prevotella buccalis*
Chlamydophila pneumoniae	*Prevotella corporis*
Chromobacterium violaceum	*Prevotella dentalis*
Chryseobacterium meningosepticum	*Prevotella denticola*
Chryseobacterium balustinum	*Prevotella disiens*
Citrobacter amalonaticus	*Prevotella enoeca*
Citrobacter braakii	*Prevotella intermedia*

Citrobacter farmeri	*Prevotella loescheii*
Citrobacter freundii	*Prevotella melaninogenica*
Citrobacter koseri	*Prevotella nigrescens*
Citrobacter rodentium	*Prevotella oralis*
Citrobacter sedlakii	*Prevotella oris*
Citrobacter werkmanii	*Prevotella ruminicola*
Citrobacter youngae	*Prevotella veroralis*
Clostridium bifermentans	*Propionibacterium avidum*
Clostridium difficile	*Propionibacterium granulosum*
Clostridium histolyticum	*Propionibacterium propionicus*

Clostridium ramosum	*Propionibacterium acnes*
Clostridium sordellii	*Proteus mirabilis*
Clostridium tertium	*Proteus penneri*
Clostridium baratii	*Proteus vulgaris*
Clostridium botulinum	*Providencia alcalifaciens*
Clostridium butyricum	*Providencia rettgeri*
Clostridium chauvoei	*Providencia stuartii*
Clostridium fallax	*Pseudomonas aeruginosa*
Clostridium novyi	*Pseudomonas alcaligenes*
Clostridium perfringens	*Pseudomonas fluorescens*
Clostridium septicum	*Pseudomonas pseudoalcaligenes*
Clostridium sporogenes	*Pseudomonas putida*
Clostridium tetani	*Pseudomonas stutzeri*
Collinsella aerofaciens	*Pseudonocardia autotrophica*
Comamonas testosteroni	*Pseudoramibacter alactolyticus*
Corynebacterium macginleyi	*Psychrobacter phenylpyruvicus*
Corynebacterium pseudodiphthericum	*Rahnella aquatilis*
Corynebacterium afermentans	*Ralstonia pickettii*
Corynebacterium amycolatum	*Rhodococcus equi*
Corynebacterium argentoratense	*Rhodococcus erythropolis*
Corynebacterium bovis	*Rhodococcus fascians*
Corynebacterium diphtheriae	*Rhodococcus rhodnii*
Corynebacterium jeikeium	*Rhodococcus rhodochrous*

Corynebacterium kutscheri	*Rickettsia africae*
Corynebacterium minutissimum	*Rickettsia akari*
Corynebacterium propinquum	*Rickettsia australis*
Corynebacterium pseudotuberculosis	*Rickettsia conorii*
Corynebacterium striatum	*Rickettsia felis*
Corynebacterium ulcerans	*Rickettsia honei*
Corynebacterium urealyticum	*Rickettsia japonica*
Corynebacterium xerosis	*Rickettsia massiliae*
Coxiella burnetii	*Rickettsia prowazekii*
Delftia acidovorans	*Rickettsia rickettsii*
Dermatophilus congolensis	*Rickettsia sibirica*
Dichelobacter nodosus	*Rickettsia typhi*
Edwardsiella hoshinae	*Rothia dentocariosa*
Edwardsiella tarda	*Ruminococcus productus*
Eggerthella lenta	*Saccharomonospora viridis*
Ehrlichia ewingii	*Saccharopolyspora rectivirgula*
Ehrlichia chaffeensis	*Salmonella choleraesuis*
Eikenella corrodens	*Salmonella enteritidis*
Enterobacter aerogenes	*Salmonella typhi*
Enterobacter amnigenus	*Salmonella typhimurium*
Enterobacter gergoviae	*Salmonella bongori*
Enterobacter sakazakii	*Sebaldella termitidis*
Enterobacter asburiae	*Selenomonas dianae*
Enterobacter cancerogenus	*Selenomonas artemidis*

Enterobacter cloacae	*Selenomonas flueggei*
Enterobacter hormaechei	*Selenomonas infelix*
Enterococcus avium	*Selenomonas noxia*
Enterococcus casseliflavus	*Serratia ficaria*
Enterococcus durans	*Serratia marcescens*
Enterococcus faecalis	*Serratia odorifera*
Enterococcus faecium	*Serratia plymuthica*
Enterococcus flavescens	*Serratia proteamaculans*
Enterococcus gallinarum	*Serratia rubidaea*
Enterococcus hirae	*Shigella boydii*

Enterococcus mundtii	*Shigella dysenteriae*
Enterococcus raffinosus	*Shigella flexneri*
Erysipelothrix rhusiopathiae	*Shigella sonnei*
Escherichia coli	*Sphingomonas paucimobilis*
Eubacterium combesii	*Spirillum minus*
Eubacterium contortum	*Staphylococcus aureus*
Eubacterium cylindroides	*Staphylococcus epidermidis*
Eubacterium moniliforme	*Staphylococcus haemolyticus*
Eubacterium multiforme	*Staphylococcus hyicus*
Eubacterium saburreum	*Staphylococcus intermedius*
Eubacterium tenue	*Staphylococcus lugdunensis*
Eubacterium brachy	*Staphylococcus saprophyticus*
Eubacterium limosum	*Staphylococcus warneri*
Eubacterium nodatum	*Stenotrophomonas maltophilia*
Eubacterium rectale	*Streptobacillus moniliformis*
Eubacterium saphenum	*Streptococcus bovis*
Eubacterium sulci	*Streptococcus milleri*
Ewingella americana	*Streptococcus sanguis*
Fibrobacter intestinalis	*Streptococcus acidominimus*
Filifactor alocis	*Streptococcus agalactiae*
Finegoldia magna	*Streptococcus anginosus*
Fluoribacter bozemanae	*Streptococcus canis*
Fluoribacter dumoffii	*Streptococcus constellatus*
Fluoribacter gormanii	*Streptococcus criceti*
Francisella tularensis	*Streptococcus equi*
Fusobacterium mortiferum	*Streptococcus gordonii*
Fusobacterium necrophorum	*Streptococcus intermedius*
Fusobacterium nucleatum	*Streptococcus mitis*
Fusobacterium periodonticum	*Streptococcus mutans*
Fusobacterium ulcerans	*Streptococcus pneumoniae*
Fusobacterium varium	*Streptococcus pyogenes*
Gardnerella vaginalis	*Streptococcus salivarius*
Gemella morbillorum	*Streptococcus sobrinus*
Gordonia rubropertincta	*Streptococcus suis*
Gordonia amarae	*Streptococcus uberis*

Gordonia bronchialis	*Sutterella wadsworthensis*
Gordonia sputi	*Suttonella indologenes*

Gordonia terrae	*Tatlockia maceachernii*
Granulicatella adjacens	*Tatlockia micdadei*
Haemophilus aphrophilus	*Tatumella ptyseos*
Haemophilus paraphrophilus	*Treponema carateum*
Haemophilus segnis	*Treponema pallidum*
Haemophilus ducreyi	*Tropheryma whippelii*
Haemophilus haemolyticus	*Tsukamurella inchonensis*
Haemophilus influenzae	*Tsukamurella paurometabola*
Haemophilus parahaemolyticus	*Tsukamurella pulmonis*
Haemophilus parainfluenzae	*Tsukamurella tyrosinosolvens*
Hafnia alvei	*Ureaplasma urealyticum*
Helicobacter cinaedi	*Veillonella atypica*
Helicobacter fennelliae	*Veillonella dispar*
Helicobacter heilmannii	*Veillonella parvula*
Helicobacter pullorum	*Vibrio hollisae*
Helicobacter pylori	*Vibrio alginolyticus*
Kingella denitrificans	*Vibrio cholerae*
Kingella kingae	*Vibrio cincinnatiensis*
Klebsiella granulomatis	*Vibrio fluvialis*
Klebsiella ornithinolytica	*Vibrio furnissii*
Klebsiella oxytoca	*Vibrio mimicus*
Klebsiella pneumoniae	*Vibrio parahaemolyticus*
Kluyvera ascorbata	*Vibrio vulnificus*
Kluyvera cryocrescens	*Wolinella succinogenes*
Lactobacillus sp.	*Yersinia bercovieri*
Legionella anisa	*Yersinia enterocolitica*
Legionella birminghamensis	*Yersinia frederiksenii*
Legionella cherrii	*Yersinia intermedia*
Legionella cincinnatiensis	*Yersinia kristensenii*
Legionella feeleii	*Yersinia mollaretii*
Legionella hackeliae	*Yersinia pestis*
Legionella jordanis	*Yersinia pseudotuberculosis*

Legionella lansingensis	*Yersinia rohdei*
Legionella longbeachae	*Yersinia ruckeri*

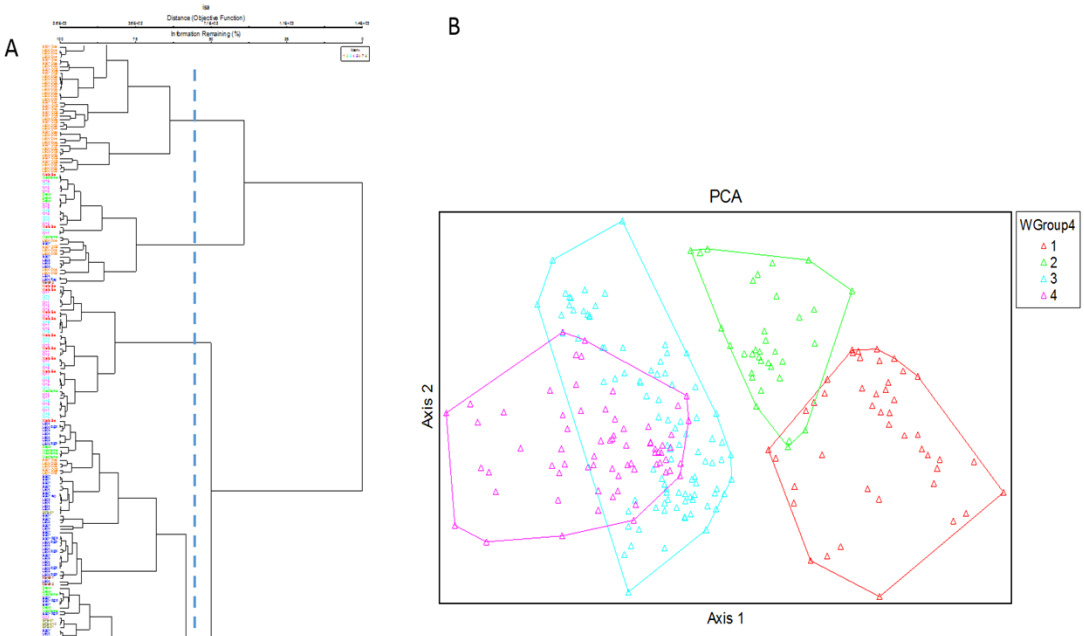
Supplementary Material 1B: List of 83 (from 583 total) fecal specific potential pathogens

Bacteroides distasonis	*Enterobacter cloacae*
Bacteroides eggerthii	*Enterobacter gergoviae*
Bacteroides forsythus	*Enterobacter hormaechei*
Bacteroides fragilis	*Enterobacter sakazakii*
Bacteroides galacturonicus	*Enterococcus avium*
Bacteroides merdae	*Enterococcus casseliflavus*
Bacteroides ovatus	*Enterococcus durans*
Bacteroides pectinophilus	*Enterococcus faecalis*
Bacteroides splanchnicus	*Enterococcus faecium*
Bacteroides stercoris	*Enterococcus flavescens*
Bacteroides thetaiotaomicron	*Enterococcus gallinarum*
Bacteroides uniformis	*Enterococcus hirae*
Bacteroides ureolyticus	*Enterococcus mundtii*
Bacteroides vulgatus	*Enterococcus raffinosus*
Campylobacter coli	*Escherichia coli*
Campylobacter concisus	*Helicobacter cinaedi*
Campylobacter curvus	*Helicobacter fennelliae*
Campylobacter fetus	*Helicobacter heilmannii*
Campylobacter gracilis	*Helicobacter pullorum*
Campylobacter jejuni	*Helicobacter pylori*
Campylobacter lari	*Klebsiella granulomatis*
Campylobacter rectus	*Klebsiella ornithinolytica*
Campylobacter sputorum	*Klebsiella oxytoca*
Campylobacter upsaliensis	*Klebsiella pneumoniae*
Citrobacter amalonaticus	*Prevotella bivia*
Citrobacter braakii	*Prevotella buccae*
Citrobacter farmeri	*Prevotella buccalis*
Citrobacter freundii	*Prevotella corporis*
Citrobacter koseri	*Prevotella dentalis*

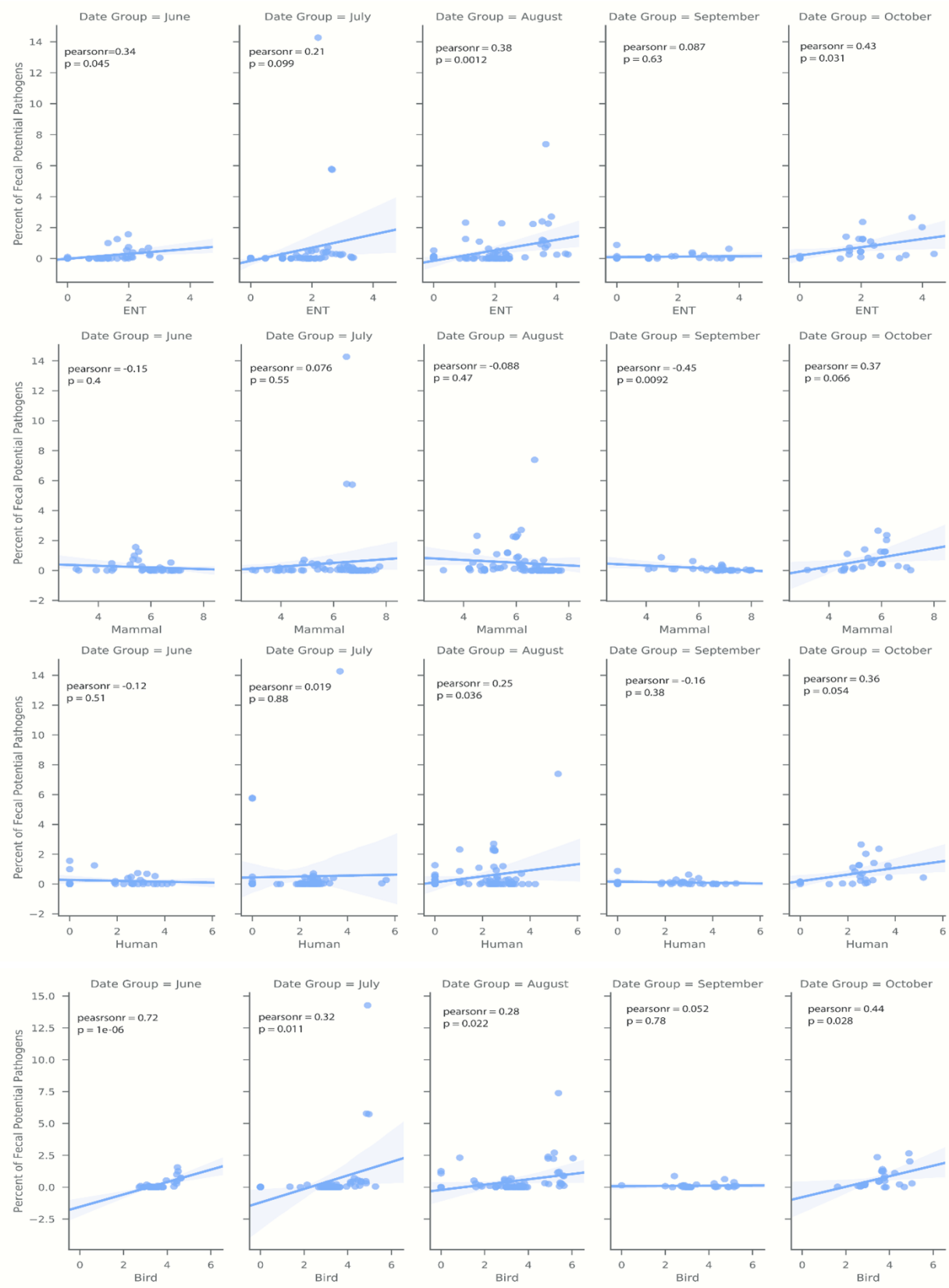
Citrobacter rodentium	*Prevotella denticola*
Citrobacter sedlakii	*Prevotella disiens*
Citrobacter werkmanii	*Prevotella enoeca*
Citrobacter youngae	*Prevotella heparinolytica*
Clostridium baratii	*Prevotella intermedia*
Clostridium bifermentans	*Prevotella loescheii*
Clostridium botulinum	*Prevotella melaninogenica*
Clostridium butyricum	*Prevotella nigrescens*
Clostridium chauvoei	*Prevotella oralis*
Clostridium difficile	*Prevotella oris*
Clostridium fallax	*Prevotella oulora*
Clostridium histolyticum	*Prevotella ruminicola*
Clostridium novyi	*Prevotella tannerae*
Clostridium perfringens	*Prevotella veroralis*
Clostridium ramosum	*Prevotella zoogleoformans*
Clostridium septicum	*Salmonella bongori*
Clostridium sordellii	*Salmonella choleraesuis*
Clostridium sporogenes	*Salmonella enteritidis*

Clostridium tertium	*Salmonella typhi*
Clostridium tetani	*Salmonella typhimurium*
Enterobacter aerogenes	*Shigella boydii*
Enterobacter amnigenus	*Shigella dysenteriae*
Enterobacter asburiae	*Shigella flexneri*
Enterobacter cancerogenus	*Shigella sonnei*

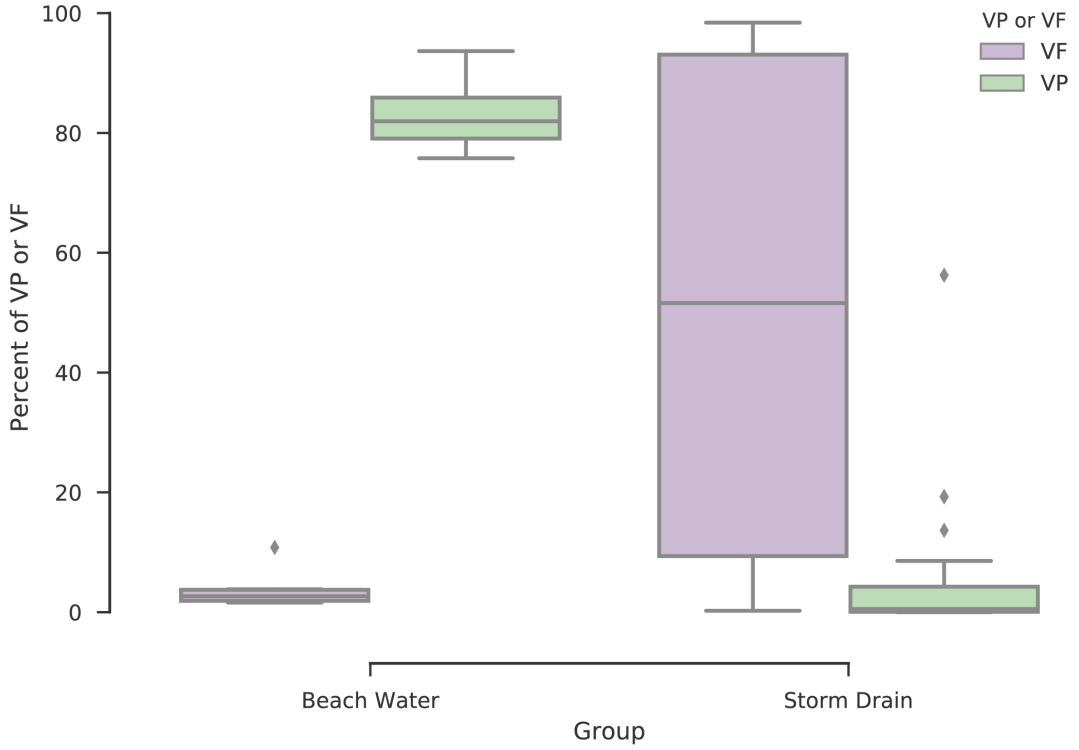
Supplementary Material 2 – Clustering. (A) Original Dendrogram from Ward's clustering (B) PCA of Clustered Groups



Supplementary Material 3: Seasonal Linear Regression for all markers and fecal specific potential pathogens



Supplementary 4. Boxplot difference between Vp and Vf in storm drains vs. beach water of outlier data (>5% potential pathogens)



References

1. Senate U. 1948. Federal Water Pollution Control Act.

2. BONDE GJ. 1966. Bacteriological Methods for Estimation of Water Pollution. Health Lab Sci 3:124–128.

3. US EPA. 1986. Ambient Water Quality Criteria for Bacteria - 1986 United States Environmental Protection Agency,Office of Water Regulations and Standards Criteria and Standards Division.

4. Cabelli VJ. 1989. Swimming-Associated Illness and Recreational Water Quality Criteria 2:13–21.

5. Cabelli VJ, Dufour AP, McCabe LJ, Levin MA. 1983. A Marine Recreational Water Quality Criterion Consistent with Indicator Concepts and Risk Analysis. J Water Pollut Control Fed 55:1306–1314.

6. Dufour AP. 1984. Bacterial indicators of recreational water quality. Can J Public Heal 75:49–56.

7. Colford JM, Wade TJ, Schiff KC, Wright CC, Griffith JF, Sandhu SK, Burns S, Sobsey M, Lovelace G, Weisberg SB. 2007. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. Epidemiology 18:27–35.

8. Soller JA, Schoen ME, Bartrand T, Ravenscroft JE, Ashbolt NJ. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. Water Res 44:4674–4691.

9. Soller JA, Schoen ME, Varghese A, Ichida AM, Boehm AB, Eftim S, Ashbolt NJ, Ravenscroft JE. 2014. Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. Water Res 66:254–264.

10. WHO. 1999. Health - Based Monitoring of Recreational Waters : the Feasibility of a New Approach.

11. Abdelzaher AM, Wright ME, Ortega C, Solo-Gabriele HM, Miller G, Elmir S, Newman X, Shih P, Alfredo Bonilla J, Bonilla TD, Palmer CJ, Scott T, Lukasik J, Harwood VJ, McQuaig S, Sinigalliano C, Gidley M, Plano LRW, Zhu X, Wang JD, Fleming LE. 2010. Presence of pathogens and indicator microbes at a non-point source subtropical recreational marine beach. Appl Environ Microbiol 76:724–732.

12. Garcia-Armisen T, Servais P. 2007. Respective contributions of point and non-point sources of E. coli and enterococci in a large urbanized watershed (the Seine river, France). J Environ Manage 82:512–518.

13. EPA. 2012. Recreational Water Quality Criteria. U S Environ Prot Agency 1–69.

14. Agency USEP, Unit W, Tech T, Consultants HE. 2011. Using Microbial Source Tracking to Support TMDL Development and Implementation Acknowledgments.

15. Anderson RJ, Woodward RL. 1969. Concepts of fecal streptococci in stream pollution. J Water Pollut Control Fed 41:R336–R352.

16. Scott TM, Rose JB, Jenkins TM, Samuel R, Lukasik J, Farrah SR. 2002. Microbial Source Tracking : Current Methodology and Future Directions Microbial Source Tracking : Current Methodology and Future Directions †. Appl Environ Microbiol 68:5796–5803.

17. Hagedorn C, Robinson SL, Filtz JR, Grubbs SM, Angier TA, Reneau RB. 1999. Determining Sources of Fecal Pollution in a Rural Virginia Watershed with Antibiotic Resistance Patterns in Fecal Streptococci 65:5522–5531.

18. SANTO-DOMINGO J, HANSEL J, MOLINA M, OSHIRO R, SHANKS OC, STELMA GN, EDGE T, GRIFFITH J, HARWOOD V, JENKINS M, LAYTON A, NAKATSU C, SADOSWKY M, STEWART J, STOECKEL D, WIGGINS B, WILBUR. J. 2004. Microbial Source Tracking Guide Document.

19. Carson CA, Shear BL, Ellersieck MR. 2001. Identification of Fecal Escherichia coli from Humans and Animals by Ribotyping Identification of Fecal *Escherichia coli* from Humans and Animals by Ribotyping. Appl Environ Mcrobiology 67:1503–1507.

20. Hartel PG, Summer JD, Hill JL, Collins J V, Entry JA, Segars WI. 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. JEnvironQual 31:1273–1278.

21. Meays CL, Broersma K, Nordin R, Mazumder A. 2004. Source tracking fecal bacteria in water : a critical review of current methods 73:71–79.

22. Griffith JF, Weisberg SB, Mcgee CD, Griffith JF. 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples 141–151.

23. Jones SH. 2011. Microbial Pathogens and Biotoxins: State of the Gulf of Maine Report.

24. Given S, Boehm AB. 2006. Regional Public Health Cost Estimates of Contaminated Coastal Waters : A Case Study of Gastroenteritis at Southern California Beaches 4851–4858.

25. Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A. 2014. Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. FEMS Microbiol Rev 38:1–40.

26. WHO. Guidelines for drinking-water quality: Fourth edition incorpoating the first addendum: Ch 4 Faecal pollution and water quality. 2017.

27. Staley C, Reckhow KH, Lukasik J, Harwood VJ. 2012. Assessment of sources of human pathogens and fecal contamination in a Florida freshwater lake. Water Res 46:5799–5812.

28. Hata a., Kitajima M, Katayama H. 2013. Occurrence and reduction of human viruses, F-specific RNA coliphage genogroups and microbial indicators at a full-scale wastewater treatment plant in Japan. J Appl Microbiol 114:545–554.

29. Cai L, Zhang T. 2013. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. Environ Sci Technol 47:5433–5441.

30. Sowah RA, Habteselassie MY, Radcliffe DE, Bauske E, Risse M. 2017. Isolating the impact of septic systems on fecal pollution in streams of suburban watersheds in Georgia , United States. Water Res 108:330–338.

31. Wright ME, Solo-gabriele HM, Elmir S, Fleming LE. 2009. Microbial load from animal feces at a recreational beach. Mar Pollut Bull 58:1649–1656.

32. Brown KI, Graham KE, Boehm AB. 2017. Risk-Based Threshold of Gull-Associated Fecal Marker Concentrations for Recreational Water.

33. Ashbolt NJ, Agency USEP, Exposure N. 2010. Assessing Pathogen Risk to Swimmers at Non-Sewage Impacted Recreational Beaches 44:2286–2291.

34. Converse RR, Kinzelman JL, Sams EA, Hudgens E, Dufour AP, Ryu H, Santo-domingo JW, Kelty CA, Shanks OC, Siefring SD, Haugland RA, Wade TJ. 2012. Dramatic Improvements in Beach Water Quality Following Gull Removal.

35. Nugent B, Gagne K, Dillingham MJ. 2007. Managing Gulls to Reduce Fecal Coliform Bacteria in a Municipal Drinking Water Source.

36. Nelson M, Jones SH, Edwards C, Ellis JC. 2008. Characterization of Escherichia coli populations from gulls , landfill trash , and wastewater using ribotyping 81:53–63.

37. Cinquepalmi V, Monno R, Fumarola L, Ventrella G. 2013. Environmental Contamination by Dog ' s Faeces : A Public Health Problem? 72–84.

38. Green HC, White KM, Kelty CA, Shanks OC. 2014. Development of Rapid Canine Fecal Source Identi fi cation PCR-Based Assays.

39. Higuita NIA, Huycke MM. 2014. Enterococcal Disease , Epidemiology , and Implications for Treatment 1–35.

40. Devriese LUCA, Ieven M, Goossens H, Vandamme P, Pot B, Hommez J, Haesebrouck F. 1996.

Presence of Vancomycin-Resistant Enterococci in Farm and Pet Animals 40:2285–2287.

41. Cleaveland S, Laurenson MK, Taylor LH. 2001. Diseases of humans and their domestic mammals : pathogen characteristics , host range and the risk of emergence.

42. Coletti C, McMillan B, Soule S. 2007. The Inside Scoop : How to Conduct a Pet Waste Outreach Campaign.

43. Clark DE. 2014. Roosting , Site Fidelity , and Food Sources of Urban Gulls In Massachusetts : Implications For Protecting Public Water Supplies.

44. Whiting-Grant K, Jones SH, Dionne M, Dillion F, Dalton C. 2004. MICROBIAL SOURCE TRACKING IN TWO SOUTHERN MAINE WATERSHEDS - MBLR Watershed.

45. Craun GF, Calderon RL, Craun MF. 2007. Outbreaks associated with recreational water in the United States 3123.

46. Icrobiol APPLENM. 2005. Concentrations of Pathogens and Indicators in Animal Feces in the Sydney Watershed 71:5929–5934.

47. Byappanahalli MN, Nevers MB, Korajkic A, Staley ZR, Harwood VJ. 2012. Enterococci in the environment. Microbiol Mol Biol Rev 76:685–706.

48. Anderson KL, Whitlock JE, Valerie J, Harwood VJ. 2005. Persistence and Differential Survival of Fecal Indicator Bacteria in Subtropical Waters and Sediments Persistence and Differential Survival of Fecal Indicator Bacteria in Subtropical Waters and Sediments. Appl Environ Microbiol 71:3041–3048.

49. Haller L, Amedegnato E, Poté J, Wildi W. 2009. Influence of Freshwater Sediment Characteristics on Persistence of Fecal Indicator Bacteria. Water Air Soil Pollut 1–11.

50. Badgley BD, Thomas FIM, Harwood VJ. 2010. The effects of submerged aquatic vegetation on the persistence of environmental populations of Enterococcus spp. Environ Microbiol 12:1271–1281.

51. Jeng HC, Sinclair R, Daniels R, Englande AJ. 2005. Survival of *Enterococci facalis* in estuarine sediments. Int J Environ Stud 62:283–291.

52. Jamieson RC, Gordon RJ, Sharples KE, Stratton GW, Madani A. 2002. Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review. Can Biosyst Eng 44:1.1-1.9.

53. Desmarais TR, Solo-gabriele HM, Carol J, Palmer CJ. 2002. Influence of Soil on Fecal Indicator

Organisms in a Tidally Influenced Subtropical Environment Influence of Soil on Fecal Indicator Organisms in a Tidally Influenced Subtropical Environment. Appl Environ Microbiol 68:1165–1172.

54. Byappanahalli MN, Roll BM, Fujioka RS. 2012. Evidence for Occurrence, Persistence, and Growth Potential of Escherichia coli and Enterococci in Hawaii^|^rsquo;s Soil Environments. Microbes Environ 27:164–170.

55. Fujioka R, Sian-Denton C, Borja M, Castro J, Morphew K. 1998. Soil: the environmental source of *Escherichia coli* and Enterococci in Guam's streams. J Appl Microbiol 85 Suppl 1:83S–89S.

56. Doran JW. 1979. Bacteriological Quality of Runoff Water from Pastureland Bacteriological Quality of Runoff Water from Pasturelandt.

57. Howell JM, Coyne MS. 1995. Fecal Bacteria in Agricultural Waters of the Bluegrass Region of Kentucky.

58. Byappanahalli M, Fujioka R, Resources W, Michigan L. 2002. Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils 27–32.

59. Anderson KL, Whitlock JE, Valerie J, Harwood VJ. 2005. Persistence and Differential Survival of Fecal Indicator Bacteria in Subtropical Waters and Sediments Persistence and Differential Survival of Fecal Indicator Bacteria in Subtropical Waters and Sediments. Appl Environ Microbiol 71:3041–3048.

60. Craig DL, Fallowfield HJ, Cromar NJ. 2002. Enumeration of faecal coliforms from recreational coastal sites: Evaluation of techniques for the separation of bacteria from sediments. J Appl Microbiol 93:557–565.

61. Ferguson DM, Moore DF, Getrich MA, Zhowandai MH. 2005. Enumeration and speciation of enterococci found in marine and intertidal sediments and coastal water in southern California. J Appl Microbiol 99:598–608.

62. Jamieson RC, Joy DM, Lee H, Kostaschuk R, Gordon RJ. 2005. Resuspension of Sediment-Associated Escherichia coli in a Natural Stream 589:581–589.

63. Brinkmeyer R, Amon RMW, Schwarz JR, Saxton T, Roberts D, Harrison S, Ellis N, Fox J, Diguardi K, Hochman M, Duan S, Stein R, Elliott C. 2015. Science of the Total Environment Distribution and persistence of Escherichia coli and Enterococci in stream bed and bank sediments from two urban streams in Houston , TX. Sci Total Environ 502:650–658.

64. Jones S, Sumner S, Landry N, Connor J. 2007. Pollution Source Tracking at New Hampshire (USA) Ocean Beaches.

65. Moore DF, Harwood VJ, Ferguson DM, Lukasik J, Hannah P, Getrich M, Brownell M. 2005. Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed 618–628.

66. Bernhard AE, Field KG. 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. Appl Environ Microbiol.

67. Bernhard AE, Field KG. 2000. Identification of Nonpoint Sources of Fecal Pollution in Coastal Waters by Using Host-Specific 16S Ribosomal DNA Genetic Markers from Fecal Anaerobes. Appl Environ Microbiol 66:1587–1594.

68. Kreader. 1995. Design and evaluation of Bacteroides DNA probes for the specific detection of human fecal pollution. Appl Environ Microbiol 61:1171–1179.

69. Fogarty L, Voytek M. 2005. Comparison of Bacteroides-Prevotella 16S rRNA genetic markers for fecal samples from different animal species. Appl Environ Microbiol 71:5999–6007.

70. Ahmed W, Goonetilleke A, Powell D, Chauhan K, Gardner T. 2009. Comparison of molecular markers to detect fresh sewage in environmental waters. Water Res 43:4908–4917.

71. Ahmed W, Goonetilleke a., Powell D, Gardner T. 2009. Evaluation of multiple sewage-associated Bacteroides PCR markers for sewage pollution tracking. Water Res 43:4872–4877.

72. Seurinck S, Defoirdt T, Verstraete W, Siciliano SD. 2005. Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. Environ Microbiol 7:249–259.

73. Converse RR, Blackwood AD, Kirs M, Griffith JF, Noble RT. 2009. Rapid QPCR-based assay for fecal Bacteroides spp. as a tool for assessing fecal contamination in recreational waters. Water Res 43:4828–4837.

74. Green HC, Haugland R a., Varma M, Millen HT, Borchardt M a., Field KG, Walters W a., Knight R, Sivaganesan M, Kelty C a., Shanks OC. 2014. Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. Appl Environ Microbiol 80:3086–3094.

75. Dick LK, Simonich MT, Field KG. 2005. Microplate Subtractive Hybridization To Enrich for Bacteroidales Genetic Markers for Fecal Source Identification. Appl Environ Microbiol 71:3179–3183.

76. Christopher D. Sinigallianoa, Jay M. Fleisherb, Maribeth L. Gidleya, Helena M. SoloGabrieleb E, Tomoyuki Shibataa, Lisa R.W. Planob F, , Samir M. Elmirb D, Wanlessa, Jakub Bartkowiaka, Rene Boiteaua, Kelly Withumb AM, Abdelzaherb, Guoqing Heb, Cristina Ortegab, Xiaofang

Zhub, Mary E. Wrightb E, Jonathan Kishb F, , Julie Hollenbeckb, Troy Scotth LCB, E. and L, Flemingb. 2010. Traditional and Molecular Analyses for Fecal Indicator Bacteria in Non-point SOurce Subtropical Recreational Marine Waters. Water Res 27:590–609.

77. Schriewer A, Goodwin KD, Sinigalliano CD, Cox AM, Wanless D, Bartkowiak J, Ebentier DL, Hanley KT, Ervin J, Deering LA, Shanks OC, Peed LA, Meijer WG, Griffith JF, SantoDomingo J, Jay JA, Holden PA, Wuertz S. 2013. Performance evaluation of canine-associated Bacteroidales assays in a multi-laboratory comparison study. Water Res 47:6909–6920.

78. Boehm AB, Van De Werfhorst LC, Griffith JF, Holden PA, Jay JA, Shanks OC, Wang D, Weisberg SB. 2013. Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. Water Res 47:6812–6828.

79. Jones SH. 2008. Environmental sources of microbial contaminants in shellfish and their public health significance 238–244.

80. Lu J, Santo Domingo JW, Lamendella R, Edge T, Hill S. 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl Environ Microbiol.

81. Sinigalliano CD, Ervin JS, Van De Werfhorst LC, Badgley BD, Ballesté E, Bartkowiak J, Boehm AB, Byappanahalli M, Goodwin KD, Gourmelon M, Griffith J, Holden P a., Jay J, Layton B, Lee C, Lee J, Meijer WG, Noble R, Raith M, Ryu H, Sadowsky MJ, Schriewer A, Wang D, Wanless D, Whitman R, Wuertz S, Santo Domingo JW. 2013. Multi-laboratory evaluations of the performance of Catellicoccus marimammalium PCR assays developed to target gull fecal sources. Water Res 47:6883–6896.

82. Jones SH, McLeod et al. (Eds). 2014. Microbial Source Tracking to Identify and Manage Sources of Fecal Contamination in Shellfish Growing Waters, Ch 1. Proceedings of the 9th International Conference of Molluscan Shellfish Safety, Sydney.

83. Jones SH, Shumway S, Rodrick G (Eds). 2009. Microbial Contamination and Shellfish Safety, Ch1. Woodhead Publishing, Cambridge, UK.

84. Green HC, Dick LK, Gilpin B, Samadpour M, Field KG. 2012. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. Appl Environ Microbiol 78:503–510.

85. Ahmed W, Harwood VJ, Nguyen K, Young S, Hamilton K, Toze S. 2016. Utility of Helicobacter spp. associated GFD markers for detecting avian fecal pollution in natural waters of two continents. Water Res 88:613–622.

86. Tan B, Ng C, Nshimyimana JP, Loh LL, Gin KY-H, Thompson JR. 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Front Microbiol 6.

87. Mclellan SL, Eren AM. 2014. Discovering new indicators of fecal pollution. Trends Microbiol 22:697–706.

88. Newton RJ, Bootsma MJ, Morrison HG, Sogin ML, McLellan SL. 2013. A Microbial Signature Approach to Identify Fecal Pollution in the Waters Off an Urbanized Coast of Lake Michigan. Microb Ecol 65:1011–1023.

89. Knights, Dan, Kucyznski, Justin, Charlson, Emily S, Zaneveld, Jesse, Mozer, Michael C, Collman RG, Bushman FD, Knight R, Kelley ST. 2013. Bayesian community-wide culture-independent microbial source tracking. Nat Methods 8:761–763.

90. Rebekah Henry, Schang C, Coutts S, Kolotelo P, Prosser T, Crosbie N, Grant T, Cottam D, O'Brien P, Deletic A, McCarthy D. 2016. Into the deep: evaluation of SourceTracker for assessment of faecal contamination of coastal waters. Water Res 93:242–253.

91. Hewitt KM, Mannino FL, Gonzalez A, Chase JH, Caporaso JG, Knight R, Kelley ST. 2013. Bacterial Diversity in Two Neonatal Intensive Care Units (NICUs) 8.

92. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. 2016. Sediments and Soils Act as Reservoirs for Taxonomic and Functional Bacterial Diversity in the Upper Mississippi River. Microb Ecol.

93. Thomsen PF, Willerslev E. 2015. Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183:4–18.

94. Barnes MA, Turner CR, Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. Conserv Genet 17:1–17.

95. Cameron TR. 2014. Particle size distribution and optimal capture of aqueous macrobial eDNA bioRxiv 1–40.

96. Martellini A, Payment P, Villemur R. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine and ovine sources in fecally contaminated surface water. Water Res 39:541—548.

97. Hussein K.R. 2016. Survival and persistence of *Bacteroides* species as faecal indicators and the recovery of 16S rRNA markers under controlled ... Environmental Microbiology.

98. Dong S, Hong P, Nguyen TH. 2014. Persistence of *Bacteroides ovatus* under simulated sunlight irradiation Persistence of Bacteroides ovatus under simulated sunlight irradiation.

99. Kim M, Wuertz S. 2015. Survival and persistence of host-associated *Bacteroidales* cells and DNA in comparison with *Escherichia coli* and Enterococcus in freshwater sediments as quantified by

PMA-qPCR and qPCR. Water Res 87:182–192.

100. Jones SH, Urquhart EA, Guimond J. Water Quality & Statistics-Based Precautionary Advisory Tool for York, Maine Beaches.

101. MHB. 2014. Maine Healthy Beaches 2014 Report to US EPA.

102. He LL, He Z. 2008. Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in 42:2563–2573.

103. Jones SH, Landry N, Hampshire N, Landry N. 2003. Tracking Bacterial Pollution Sources in Hampton Harbor.

104. Jones SH. 2007. Microbial Pollution Source Tracking at New Castle Beach.

105. Jones SH, Rothenheber D. 2015. Microbial Source Tracking at York Beach Storm Drains 3824.

106. Jones SH, Rothenheber D. 2017. Mammal and Human Fecal Contamination Source Identification in South Portland , Maine.

107. Boehm AB, Soller JA, Shanks OC. 2015. Human-Associated Fecal Quantitative Polymerase Chain Reaction Measurements and Simulated Risk of Gastrointestinal Illness in Recreational Waters Contaminated with Raw Sewage. Environ Sci Technol Lett 2:270–275.

108. Dorfmam M, Haren A. 2013. Testing the Waters. NRDC.

109. Whiting-Grant K, Dillion F, Dalton C. 2003. MICROBIAL SOURCE TRACKING IN TWO SOUTHERN MAINE WATERSHEDS -Webhannet.

110. Boehm AB, Sassoubre LM. 2014. Enterococci as Indicators of Environmental Fecal Contamination. Enterococci From Commensals to Lead Causes Drug Resist Infect 1–19.

111. Reischer GH, Kavka GG, Kasper DC, Winter C, Mach RL, Farnleitner AH. 2008. Applicability of DNA based quantitative microbial tracking (QMST) evaluated on a large scale in the Danube River and its important tributaries. Large Rivers 18:117–125.

112. Ebentier DL, Hanley KT, Cao Y, Badgley BD, Boehm AB, Ervin JS, Goodwin KD, Gourmelon M, Griffith JF, Holden P a., Kelty C a., Lozach S, McGee C, Peed L a., Raith M, Ryu H, Sadowsky MJ, Scott E a., Domingo JS, Schriewer A, Sinigalliano CD, Shanks OC, Van De Werfhorst LC, Wang D, Wuertz S, Jay J a. 2013. Evaluation of the repeatability and reproducibility of a suite of qPCR-based microbial source tracking methods. Water Res 47:6839–6848.

113. Fremaux B, Gritzfeld J, Boa T, Yost CK. 2009. Evaluation of host-specific *Bacteroidales* 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. Water Res 43:4838–4849.

114. Schriewer A, Miller WA, Byrne BA, Miller MA, Oates S, Conrad PA, Hardin D, Yang HH, Chouicha N, Melli A, Jessup D, Dominik C, Wuertz S. 2010. Presence of *Bacteroidales* as a predictor of pathogens in surface waters of the central California coast. Appl Environ Microbiol 76:5802–5814.

115. Flood C, Ufnar J, Wang S, Johnson J, Carr M, Ellender R. 2011. Lack of correlation between enterococcal counts and the presence of human specific fecal markers in Mississippi creek and coastal waters. Water Res 45:872–878.

116. Shibata T, Solo-Gabriele HM, Sinigalliano CD, Gidley ML, Plano LRW, Fleisher JM, Wang JD, Elmir SM, He G, Wright ME, Abdelzaher AM, Ortega C, Wanless D, Garza AC, Kish J, Scott T, Hollenbeck J, Backer LC, Fleming LE. 2010. Evaluation of Conventional and Alternative Monitoring Methods for a Recreation Marine Beach with Non-point Source of Fecal Contamination. Environ Sci Technol 44:8175–8181.

117. Santoro AE, Boehm AB. 2007. Frequent occurrence of the human-specific Bacteroides fecal marker at an open coast marine beach: Relationship to waves, tides and traditional indicators. Environ Microbiol 9:2038–2049.

118. Niewolak S. 1998. Total viable count and concentration of enteric bacteria in bottom sediments from the Czarna Hancza River , Northeast Poland. Polish J Environ Stud 7:295–306.

119. Hathaway JM, Hunt WF. 2011. Evaluation of First Flush for Indicator Bacteria and Total Suspended Solids in Urban Stormwater Runoff. Water, Air, {&} Soil Pollut 217:135–147.

120. Oshiro RK. 2002. Method 1600: Enterococci in Water by Membrane Filtration Using membrane-Enterococcus Indoxyl- β -D-Glucoside Agar (mEI)Environmetal Protection Agency.

121. Crump BC, Baross JA, Simenstad CA. 1998. Dominance of particle-attached bacteria in the Columbia River estuary, USA. Aquat Microb Ecol 14:7–18.

122. The United States Environmental Protection Agency E. 1971. Residue, Non-Filterable (Gravimetri, Dried at 103-105C) 3.

123. Sauer EP, Vandewalle JL, Bootsma MJ, Mclellan SL. 2011. Detection of the human specific Bacteroides genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. Water Res 45:4081–4091.

124. Cloutier DD, Alm EW, McLellan SL. 2015. The influence of land-use, nutrients, and geography on microbial communities and fecal indicator abundance at Lake Michigan beaches. Appl Environ

Microbiol AEM.00233-15.

125. Harwood VJ, Brownell M, Wang S, Lepo J, Ellender RD, Ajidahun A, Hellein KN, Kennedy E, Ye X, Flood C. 2009. Validation and field testing of library-independent microbial source tracking methods in the Gulf of Mexico. Water Res 43:4812–4819.

126. Shanks OC, Nietch C, Simonich M, Younger M, Reynolds D, Field KG. 2006. Basin-wide analysis of the dynamics of fecal contamination and fecal source identification in Tillamook Bay, Oregon. Appl Environ Microbiol.

127. Jokinen CC, Schreier H, Mauro W, Taboada E, Isaac-Renton JL, Topp E, Edge T, Thomas JE, Gannon VPJ. 2010. The occurrence and sources of *Campylobacter* spp., *Salmonella enterica* and *Escherichia coli* O157:H7 in the Salmon River, British Columbia, Canada. J Water Health 8:374–386.

128. Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. 2006. Development of Bacteroides 16S rRNA gene taqman-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol 72:4214–4224.

129. Mayer RE, Vierheilig J, Egle L, Reischer GH, Saracevic E, Mach RL, Kirschner AKT, Zessner M, Sommer R, Farnleitner AH. 2015. Automated Sampling Procedures Supported by High Persistence of Bacterial Fecal Indicators and *Bacteroidetes* Genetic Microbial Source Tracking Markers in Municipal Wastewater during Short-Term Storage at 5°C. Appl Environ Microbiol 81:5134–5143.

130. Haugland RA, Varma M, Sivaganesan M, Kelty C, Peed L, Shanks OC. 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected *Bacteroidales* species and human fecal waste by qPCR. Syst Appl Microbiol.

131. Nordstrom JL, Vickery MCL, Blackstone GM, Murray SL, DePaola A. 2007. Development of a multiplex real-time PCR assay with an internal amplification control for the detection of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters. Appl Environ Microbiol 73:5840–5847.

132. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108 Suppl:4516–22.

133. Caporaso JG, Lauber CL, Walters W a, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert J a, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624.

134. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez Pena A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turn KR. 2010.

QIIME allows analysis of high-throughput community sequencing data. Nat Methods 5:335–336.

135. Bolger AM, Lohse M, Usadel B. 2014. Genome analysis Trimmomatic : a flexible trimmer for Illumina sequence data 30:2114–2120.

136. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072.

137. Helland IS. 1990. Partial Least Squares Regression and Statistical Models 17:97–114.

138. Characklis GW, Dilts MJ, Simmons OD, Likirdopulos CA, Krometis LAH, Sobsey MD. 2005. Microbial partitioning to settleable particles in stormwater. Water Res 39:1773–1782.

139. Mote BL, Turner JW, Lipp EK. 2012. Persistence and growth of the fecal indicator bacteria enterococci in detritus and natural estuarine plankton communities. Appl Environ Microbiol 78:2569–2577.

140. Talke SA, Swart HE De, Schuttelaars HM. 2009. Feedback between residual circulations and sediment distribution in highly turbid estuaries : An analytical model 29:119–135.

141. Schinner T, Letzner A, Liedtke S, Castro FD, Eydelnant IA, Tufenkji N. 2010. Transport of selected bacterial pathogens in agricultural soil and quartz sand. Water Res 44:1182–1192.

142. Bushon RN, Brady AMG, Christensen ED, Stelzer EA. 2017. Multi-Year Microbial Source Tracking Study Characterizing Fecal Contamination in an Urban Watershed.

143. Wong M, Kumar L, Jenkins TM, Xagoraraki I, Phanikumar MS, Rose JB. 2009. Evaluation of public health risks at recreational beaches in Lake Michigan via detection of enteric viruses and a human-specific bacteriological marker. Water Res 43:1137–1149.

144. Meybeck M, Laroche L, Dürr HH, Syvitski JPM. 2003. Global variability of daily total suspended solids and their fluxes in rivers. Glob Planet Change 39:65–93.

145. Orton PM, Kineke GC. 2001. Comparing Calculated and Observed Vertical Suspended-Sediment Distributions from a Hudson River Estuary Turbidity Maximum. Estuar Coast Shelf Sci 52:401–410.

146. Thoe W, Gold M, Griesbach A, Grimmer M, Taggart ML, Boehm AB. 2014. Predicting water quality at Santa Monica Beach: Evaluation of five different models for public notification of unsafe swimming conditions. Water Res 67:105–117.

147. Brooks WR, Fienen MN, Corsi SR. 2013. Partial least squares for efficient models of fecal

indicator bacteria on Great Lakes beaches. J Environ Manage 114:470–475.

148. Carrascal LM, Galván I, Gordo O. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118:681–690.

149. Viau EJ, Goodwin KD, Yamahara KM, Layton BA, Sassoubre LM, Burns SL, Tong HI, Wong SHC, Lu Y, Boehm AB. 2011. Bacterial pathogens in Hawaiian coastal streams-Associations with fecal indicators, land cover, and water quality. Water Res 45:3279–3290.

150. Savichtcheva O, Okayama N, Okabe S. 2007. Relationships between *Bacteroides* 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. Water Res 41:3615–3628.

151. Li B, Ju F, Cai L, Zhang T. 2015. Profile and Fate of Bacterial Pathogens in Sewage Treatment Plants Revealed by High-Throughput Metagenomic Approach. Environ Sci Technol 49:10492–10502.

152. Ibekwe AM, Leddy M, Murinda SE. 2013. Potential human pathogenic bacteria in a mixed urban watershed as revealed by pyrosequencing. PLoS One 8:e79490.

153. Cao Y, Van De Werfhorst LC, Dubinsky E a., Badgley BD, Sadowsky MJ, Andersen GL, Griffith JF, Holden P a. 2013. Evaluation of molecular community analysis methods for discerning fecal sources and human waste. Water Res 47:6862–6872.

154. Tyrrel SF, Quinton JN. 2003. Overland flow transport of pathogens from agricultural land receiving faecal wastes. J Appl Microbiol 94 Suppl:87S–93S.

155. Marti R, Gannon VPJ, Jokinen C, Lanthier M, Lapen DR, Neumann NF, Ruecker NJ, Scott A, Wilkes G, Zhang Y, Topp E. 2013. Quantitative multi-year elucidation of fecal sources of waterborne pathogen contamination in the South Nation River basin using Bacteroidales microbial source tracking markers. Water Res 47:2315–2324.

156. Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, Rubin EM, Jansson JK. 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480:368–371.

157. Vierheilig J, Savio D, Farnleitner AH, Reischer GH, Ley RE, Mach RL, Farnleitner AH, Reischer GH. 2015. Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. Water Sci Technol 72:1962–1972.

158. Urquhart EA, Jones SH, Yu JW, Schuster BM, Marcinkiewicz AL, Whistler CA, Cooper VS. 2016. Environmental Conditions Associated with Elevated *Vibrio parahaemolyticus* Concentrations in Great Bay Estuary, New Hampshire. PLoS One 11:e0155018.

159. United States Census Bureau. 2010 Census. U.S. Census Bureau. 2010. http://www.census.gov/2010census/data/

160. Barnes KB. 2001. IMPERVIOUS SURFACES AND THE QUALITY OF NATURAL AND BUILT ENVIRONMENTS. Technical Report.

161. Skelly C, Weinstein P. 2003. Review Pathogen Survival Trajectories : An Eco-Environmental Approach to the Modeling of Human Campylobacteriosis Ecology 111:19–28.

162. Korajkic A, Mcminn BR, Shanks OC, Sivaganesan M, Fout GS, Ashbolt NJ. 2014. Biotic Interactions and Sunlight Affect Persistence of Fecal Indicator Bacteria and Microbial Source Tracking Genetic Markers in the Upper Mississippi River 80:3952–3961.

163. Jang J, Hur H, Sadowsky MJ, Byappanahalli MN, Yan T, Ishii S. 2017. Environmental *Escherichia coli* : ecology and public health implications — a review 1–12.

164. Gordon B, Callan P, Vickers C. 2008. WHO guidelines for drinking-water quality.WHO chronicle.

165. Soller JA, Bartrand T, Ashbolt NJ, Ravenscroft J, Wade TJ. 2010. Estimating the primary etiologic agents in recreational freshwaters impacted by human sources of faecal contamination. Water Res 44:4736–4747.

166. Marion JW, Lee J, Lemeshow S, Buckley TJ. 2010. Association of gastrointestinal illness and recreational water exposure at an inland U . S . beach. Water Res 44:4796–4804.

167. Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, Buckles EL, Liou S-R, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HLT, Donnenberg MS, Blattner FR. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A 99:17020–17024.

168. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821.